Standard ECMA-121
2nd Edition - December 2000

ECMA

Standardizing Information and Communication Systems

8-Bit Single-Byte Coded Graphic Character sets: Latin/Hebrew Alphabet

.

Standard ECMA-121

2nd Edition - December 2000

ECMA

Standardizing Information and Communication Systems

8-Bit Single-Byte Coded Graphic Character sets: Latin/Hebrew Alphabet

Brief History

The adoption of Standard ECMA-6 (ISO 646) in 1965 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 have been in wide use for quite some time.

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

– they did not provide all graphic characters which may be needed,

– for some characters, specially for accented letters, it was necessary to resort to BACKSPACE sequences, which created problems when processing data containing such composite characters,

– interchange among different versions was 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, provided a character set covering the requirements of most languages based on the Latin alphabet. This character set, although well suited for text communication, was difficult to use for processing as some graphic characters were 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 (which has become ISO/IEC JTC 1/SC2 in 1987) a proposal for such a coded character set. At its meeting of April 1984 SC2 decided to propose 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/IEC 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 ECMA-121 for a Latin/Hebrew coded graphic set.

   This 2nd Edition has been developed to keep it fully aligned with the new edition of ISO/IEC 8859-6.

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. These four Latin Alphabets have been submitted to ISO/IEC and JTC 1 and have become Parts 1 to 4 of ISO/IEC 8859.

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 Arabic, Cyrillic, and Greek scripts. These ECMA Standards ECMA-113, ECMA-114, and ECMA-118, resp., have become Parts 5 to 7, resp., of ISO/IEC 8859.

iv. Latin Alphabets No. 5 and No. 6 have been published as ECMA-128 and ECMA-144, resp. They have become Parts 9 and 10, resp., of ISO/IEC 8859.

This ECMA Standard has been adopted as 2nd edition of Standard ECMA-121 by the ECMA General Assembly of December 2000.

Table of contents

1   Scope

This ECMA Standard specifies a set of 155 coded graphic characters identified as the Latin/Hebrew alphabet.

This set of coded graphic characters is intended for use in data and text processing applications and also for information interchange. The set contains graphic characters used for general purpose applications in typical office environments in at least the following languages:

English, Hebrew and Latin.

It is not intended for pointed Hebrew.

This set of coded graphic characters may be regarded as a version of an 8-bit code according to Standard ECMA-35 or Standard ECMA-43 at level 1.

This ECMA Standard may not be used with any other ECMA Standards for 8-bit single-byte coded graphic character sets. If coded characters from more than one ECMA Standard are to be used together, by means of code extension techniques, the equivalent coded character sets from ISO/IEC 10367 should be used instead within a version of Standard ECMA-43 at level 2 or level 3.

The coded characters in this set may be used in conjunction with coded control functions selected from ECMA-48. However, control functions are not used to create composite graphic symbols from two or more graphic characters (see clause 6).

NOTE

This ECMA Standard is not intended for use with Telematic services defined by ITU-T. If information coded according to this ECMA Standard is to be transferred to such services, it will have to conform to the requirements of those services at the access-point.

2   Conformance

2.1   Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with this ECMA Standard if all the coded representations of graphic characters within that CC-data-element conform to the requirements of clause 6.

2.2   Conformance of devices

A device is in conformance with this ECMA Standard if it conforms to the requirements of 2.2.1, and either or both of 2.2.2 and 2.2.3. A claim of conformance shall identify the document which contains the description specified in 2.2.1.

2.2.1   Device description

A device that conforms to this ECMA Standard shall be subject of a description that identifies the means by which the user may supply characters to the device, or may recognize them when they are made available to him, as specified respectively in 2.2.2 and 2.2.3.

2.2.2   Originating devices

An originating device shall allow its user to supply any sequence of characters from those specified in clause 6, and shall be capable of transmitting their coded representations within a CC-data-element.

2.2.3   Receiving devices

A receiving device shall be capable of receiving and interpreting any coded representations of characters that are within a CC-data-element, and that conform to clause 6, and shall make the corresponding characters available to its user in such a way that the user can identify them from among those specified there, and can distinguish them from each other.

3   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 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-128   8-Bit Single-Byte Coded Graphic Character Sets - Latin alphabet No. 5

ECMA-144   8-Bit Single-Byte Coded Graphic Character Sets - Latin Alphabet No. 6

4   Definitions

For the purpose of this Standard the following definitions apply.

4.1   bi-directional text

A text which may contain strings of characters with left-to-right and right-to-left directions.

4.2   bit combination

An ordered set of bits used for the representation of characters.

4.3   byte

A bit string that is operated upon as a unit.

4.4   character

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

4.5   code table

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

4.6   coded character set; code

A set of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their bit combinations.

4.7   coded-character-data-element (CC-data-element)

An element of interchanged information that is specified to consist of a sequence of coded representations of characters, in accordance with one or more identified standards for coded character sets.

4.8   directional character properties

A set of mutually exclusive properties which may qualify the members of a character set. These properties are used by algorithms which transform text from processing sequence into presentation sequence. Examples of values for directional character properties are "right-to-left", "left-to-right", "digit", "numeric separator", "neutral".

4.9   graphic character

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

4.10   graphic symbol

A visual representation of a graphic character or of a control function.

4.11   implicit directionality

A text presentation method in which the direction is determined by an algorithm. The algorithm is based on the directional character properties of the character, its position relative to the preceding and following character and to the primary direction.

4.12   left-to-right character

A character specific to a script written from left to right like the Latin script or the Greek script. Typical examples are the letters A to Z.

4.13   position

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

4.14   right-to-left character

A character specific to a script written from right to left like the Arabic script or the Hebrew script. Typical examples are the letters of the Hebrew alphabet.

5   Notation, code table and names

5.1   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 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.

The bit combinations are also identified by notations of the form hk, where h and k are numbers in the range 0 to F in hexadecimal notation. The number h is the same as the number xx described above, and the number k the same as the number yy described above.

5.2   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. In hexadecimal notation the columns and the rows are numbered 0 to F.

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 column and row numbers are shown at the top and left edges of the table, respectively. The code table positions are also identified by notations of the form hk, where h is the column number and k is the row number in hexadecimal notation. The column and row numbers are shown at the bottom and right edges of the table, respectively.

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, or of the form hk, is the same as that of the corresponding bit combination.

5.3   Names and meanings.

This ECMA Standard assigns a unique name and a unique identifier to each graphic character. These names and identifiers have been taken from ISO/IEC 10646-1. This ECMA Standard also specifies an acronym for each of the characters SPACE, NO-BREAK SPACE, SOFT HYPHEN, LEFT-TO-RIGHT MARK and RIGHT-TO-LEFT-MARK. For acronyms only Latin capital letters A to Z are used. It is intended that the acronyms be retained in all translations of the text.

Except for SPACE (SP), NO-BREAK SPACE (NBSP), and SOFT HYPHEN (SHY), LEFT-TO-RIGHT MARK (LRM) and RIGHT-TO-LEFT MARK (RLM), this ECMA Standard does not define and does not restrict the meanings of graphic characters.

This ECMA Standard specifies a graphic symbol for each graphic character. This symbol is shown in the corresponding position of the code table. However, this Standard does not specify a particular style or font design for imaging graphic characters.

5.3.1   SPACE (SP)

A graphic character the visual representation of which consists of the absence of a graphic symbol.

5.3.2   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.

5.3.3   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 has been established within a word.

5.3.4   LEFT-TO-RIGHT MARK (LRM)

A graphic character the visual representation of which consists of the absence of a graphic symbol, which acts like a left-to-right character in a bi-directional text (such as LATIN SMALL LETTER A).

5.3.5   RIGHT-TO-LEFT MARK (RLM)

A graphic character the visual representation of which consists of the absence of a graphic symbol, which acts like a right-to-left character in a bi-directional text (such as HEBREW LETTER ALEF).

6   Specification of the coded character set

This ECMA Standard specifies 155 characters allocated to the bit combinations of the code table (table 2).

Control functions, such as BACKSPACE or CARRIAGE RETURN, shall not be used to create composite graphic symbols, which are made up from the graphic representations of two or more characters.

6.1   Characters of the set and their coded representation

See table 1.

Table 1 - Character set, coded representation

Bit combina-tion	Hex	Identifier	Name
02/00	20	U+0020	SPACE
02/01	21	U+0021	EXCLAMATION MARK
02/02	22	U+0022	QUOTATION MARK
02/03	23	U+0023	NUMBER SIGN
02/04	24	U+0024	DOLLAR SIGN
02/05	25	U+0025	PERCENT SIGN
02/06	26	U+0026	AMPERSAND
02/07	27	U+0027	APOSTROPHE
02/08	28	U+0028	LEFT PARENTHESIS
02/09	29	U+0029	RIGHT PARENTHESIS
02/10	2A	U+002A	ASTERISK
02/11	2B	U+002B	PLUS SIGN
02/12	2C	U+002C	COMMA
02/13	2D	U+002D	HYPHEN-MINUS
02/14	2E	U+002E	FULL STOP
02/15	2F	U+002F	SOLIDUS
03/00	30	U+0030	DIGIT ZERO
03/01	31	U+0031	DIGIT ONE
03/02	32	U+0032	DIGIT TWO
03/03	33	U+0033	DIGIT THREE
03/04	34	U+0034	DIGIT FOUR
03/05	35	U+0035	DIGIT FIVE
03/06	36	U+0036	DIGIT SIX
03/07	37	U+0037	DIGIT SEVEN
03/08	38	U+0038	DIGIT EIGHT
03/09	39	U+0039	DIGIT NINE
03/10	3A	U+003A	COLON
03/11	3B	U+003B	SEMICOLON
03/12	3C	U+003C	LESS-THAN SIGN
03/13	3D	U+003D	EQUALS SIGN
03/14	3E	U+003E	GREATER-THAN SIGN
03/15	3F	U+003F	QUESTION MARK
04/00	40	U+0040	COMMERCIAL AT
04/01	41	U+0041	LATIN CAPITAL LETTER A
04/02	42	U+0042	LATIN CAPITAL LETTER B
04/03	43	U+0043	LATIN CAPITAL LETTER C
04/04	44	U+0044	LATIN CAPITAL LETTER D
04/05	45	U+0045	LATIN CAPITAL LETTER E
04/06	46	U+0046	LATIN CAPITAL LETTER F
04/07	47	U+0047	LATIN CAPITAL LETTER G
04/08	48	U+0048	LATIN CAPITAL LETTER H
04/09	49	U+0049	LATIN CAPITAL LETTER I
04/10	4A	U+004A	LATIN CAPITAL LETTER J
04/11	4B	U+004B	LATIN CAPITAL LETTER K
04/12	4C	U+004C	LATIN CAPITAL LETTER L
04/13	4D	U+004D	LATIN CAPITAL LETTER M
04/14	4E	U+004E	LATIN CAPITAL LETTER N
04/15	4F	U+004F	LATIN CAPITAL LETTER O
05/00	50	U+0050	LATIN CAPITAL LETTER P
05/01	51	U+0051	LATIN CAPITAL LETTER Q

Bit combina-tion	Hex	Identifier	Name
05/02	52	U+0052	LATIN CAPITAL LETTER R
05/03	53	U+0053	LATIN CAPITAL LETTER S
05/04	54	U+0054	LATIN CAPITAL LETTER T
05/05	55	U+0055	LATIN CAPITAL LETTER U
05/06	56	U+0056	LATIN CAPITAL LETTER V
05/07	57	U+0057	LATIN CAPITAL LETTER W
05/08	58	U+0058	LATIN CAPITAL LETTER X
05/09	59	U+0059	LATIN CAPITAL LETTER Y
05/10	5A	U+005A	LATIN CAPITAL LETTER Z
05/11	5B	U+005B	LEFT SQUARE BRACKET
05/12	5C	U+005C	REVERSE SOLIDUS
05/13	5D	U+005D	RIGHT SQUARE BRACKET
05/14	5E	U+005E	CIRCUMFLEX ACCENT
05/15	5F	U+005F	LOW LINE
06/00	60	U+0060	GRAVE ACCENT
06/01	61	U+0061	LATIN SMALL LETTER A
06/02	62	U+0062	LATIN SMALL LETTER B
06/03	63	U+0063	LATIN SMALL LETTER C
06/04	64	U+0064	LATIN SMALL LETTER D
06/05	65	U+0065	LATIN SMALL LETTER E
06/06	66	U+0066	LATIN SMALL LETTER F
06/07	67	U+0067	LATIN SMALL LETTER G
06/08	68	U+0068	LATIN SMALL LETTER H
06/09	69	U+0069	LATIN SMALL LETTER I
06/10	6A	U+006A	LATIN SMALL LETTER J
06/11	6B	U+006B	LATIN SMALL LETTER K
06/12	6C	U+006C	LATIN SMALL LETTER L
06/13	6D	U+006D	LATIN SMALL LETTER M
06/14	6E	U+006E	LATIN SMALL LETTER N
06/15	6F	U+006F	LATIN SMALL LETTER O
07/00	70	U+0070	LATIN SMALL LETTER P
07/01	71	U+0071	LATIN SMALL LETTER Q
07/02	72	U+0072	LATIN SMALL LETTER R
07/03	73	U+0073	LATIN SMALL LETTER S
07/04	74	U+0074	LATIN SMALL LETTER T
07/05	75	U+0075	LATIN SMALL LETTER U
07/06	76	U+0076	LATIN SMALL LETTER V
07/07	77	U+0077	LATIN SMALL LETTER W
07/08	78	U+0078	LATIN SMALL LETTER X
07/09	79	U+007A	LATIN SMALL LETTER Y
07/10	7A	U+007A	LATIN SMALL LETTER Z
07/11	7B	U+007B	LEFT CURLY BRACKET
07/12	7C	U+007C	VERTICAL LINE
07/13	7D	U+007D	RIGHT CURLY BRACKET
07/14	7E	U+007E	TILDE
10/00	A0	U+00A0	NO-BREAK SPACE
10/01	A1		*(This position shall not be used)*
10/02	A2	U+00A2	CENT SIGN
10/03	A3	U+00A3	POUND SIGN
10/04	A4	U+00A4	CURRENCY SIGN
10/05	A5	U+00A5	YEN SIGN
10/06	A6	U+00A6	BROKEN BAR

Bit combina-tion	Hex	Identifier	Name
10/07	A7	U+00A7	SECTION SIGN
10/08	A8	U+00A8	DIARESIS
10/09	A9	U+00A9	COPYRIGHT SIGN
10/10	AA	U+00D7	MULTIPLICATION SIGN
10/11	AB	U+00AB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
10/12	AC	U+00AC	NOT SIGN
10/13	AD	U+00AD	SOFT HYPHEN
10/14	AE	U+00AE	REGISTERED SIGN
10/15	AF	U+00AF	MACRON
11/00	B0	U+00B0	DEGREE SIGN
11/01	B1	U+00B1	PLUS-MINUS SIGN
11/02	B2	U+00B2	SUPERSCRIPT TWO
11/03	B3	U+00B3	SUPERSCRIPT THREE
11/04	B4	U+00B4	ACUTE ACCENT
11/05	B5	U+00B5	MICRO SIGN
11/06	B6	U+00B6	PILCROW SIGN
11/07	B7	U+00B7	MIDDLE DOT
11/08	B8	U+00B8	CEDILLA
11/09	B9	U+00B9	SUPERSCRIPT ONE
11/10	BA	U+00F7	DIVISION SIGN
11/11	BB	U+00BB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
11/12	BC	U+00BC	VULGAR FRACTION ONE QUARTER
11/13	BD	U+00BD	VULGAR FRACTION ONE HALF
11/14	BE	U+00BE	VULGAR FRACTION THREE QUARTERS
11/15	BF		*(This position shall not be used)*
12/00	C0		*(This position shall not be used)*
12/01	C1		*(This position shall not be used)*
12/02	C2		*(This position shall not be used)*
12/03	C3		*(This position shall not be used)*
12/04	C4		*(This position shall not be used)*
12/05	C5		*(This position shall not be used)*
12/06	C6		*(This position shall not be used)*
12/07	C7		*(This position shall not be used)*
12/08	C8		*(This position shall not be used)*
12/09	C9		*(This position shall not be used)*
12/10	CA		*(This position shall not be used)*
12/11	CB		*(This position shall not be used)*
12/12	CC		*(This position shall not be used)*
12/13	CD		*(This position shall not be used)*
12/14	CE		*(This position shall not be used)*
12/15	CF		*(This position shall not be used)*
13/00	D0		*(This position shall not be used)*
13/01	D1		*(This position shall not be used)*
13/02	D2		*(This position shall not be used)*
13/03	D3		*(This position shall not be used)*
13/04	D4		*(This position shall not be used)*
13/05	D5		*(This position shall not be used)*
13/06	D6		*(This position shall not be used)*
13/07	D7		*(This position shall not be used)*
13/08	D8		*(This position shall not be used)*
13/09	D9		*(This position shall not be used)*
13/10	DA		*(This position shall not be used)*
13/11	DB		*(This position shall not be used)*

Bit combina-tion	Hex	Identifier	Name
13/12	DC		*(This position shall not be used)*
13/13	DD		*(This position shall not be used)*
13/14	DE		*(This position shall not be used)*
13/15	DF	U+2017	DOUBLE LOW LINE
14/00	E0	U+05D0	HEBREW LETTER ALEF
14/01	E1	U+05D1	HEBREW LETER BET
14/02	E2	U+05D2	HEBREW LETTER GIMEL
14/03	E3	U+05D3	HEBREW LETTER DALET
14/04	E4	U+05D4	HEBREW LETTER HE
14/05	E5	U+05D5	HEBREW LETTER VAV
14/06	E6	U+05D6	HEBREW LETTER ZAYIN
14/07	E7	U+05D7	HEBREW LETTER HET
14/08	E8	U+05D7	HEBREW LETTER TET
14/09	E9	U+05D9	HEBREW LETTER YOD
14/10	EA	U+05DA	HEBREW LETTER FINAL KAF
14/11	EB	U+05DB	HEBREW LETTER KAF
14/12	EC	U+05DC	HEBREW LETTER LAMED
14/13	ED	U+05DD	HEBREW LETTER FINAL MEM
14/14	EE	U+05DE	HEBREW LETTER MEM
14/15	EF	U+05DF	HEBREW LETTER FINAL NUN
15/00	F0	U+05E0	HEBREW LETTER NUN
15/01	F1	U+05E1	HEBREW LETTER SAMEKH
15/02	F2	U+05E2	HEBREW LETTER AYIN
15/03	F3	U+05E3	HEBREW LETTER FINAL PE
15/04	F4	U+05E4	HEBREW LETTER PE
15/05	F5	U+05E5	HEBREW LETTER FINAL TSADI
15/06	F6	U+05E6	HEBREW LETTER TSADI
15/07	F7	U+05E7	HEBREW LETTER QOF
15/08	F8	U+05E8	HEBREW LETTER RESH
15/09	F9	U+05E9	HEBREW LETTER SHIN
15/10	FA	U+05EA	HEBREW LETTER TAV
15/11	FB		*(This position shall not be used)*
15/12	FB		*(This position shall not be used)*
15/13	FD	U+200E	LEFT-TO-RIGHT MARK
15/14	FE	U+200F	RIGHT-TO-LEFT MARK
15/15	FF		*(This position shall not be used)*

6.2   Code table

For each character in the set the code table (table 2) shows a graphic symbol at the position in the code table corresponding to the bit combination specified in table 1.

The shaded positions in the code table correspond to bit combinations that do not represent graphic characters. Their use is outside the scope of this ECMA Standard; it is specified in other ECMA Standards, for example ECMA-48.

The positions in the code table that are shown with cross-hatching correspond to bit combinations in table 1 having the entry "This position shall not be used".

Table 2 - Code table of Latin/Hebrew alphabet

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	hex
0	0	0	0	00			SP	0	@	P	`	p			NBSP	°	▨	▨	א	נ	0
0	0	0	1	01			!	1	A	Q	a	q			▨	±	▨	▨	ב	ס	1
0	0	1	0	02			"	2	B	R	b	r			¢	²	▨	▨	ג	ע	2
0	0	1	1	03			#	3	C	S	c	s			£	³	▨	▨	ד	ף	3
0	1	0	0	04			$	4	D	T	d	t			¤	´	▨	▨	ה	פ	4
0	1	0	1	05			%	5	E	U	e	u			¥	µ	▨	▨	ו	ץ	5
0	1	1	0	06			&	6	F	V	f	v			¦	¶	▨	▨	ז	צ	6
0	1	1	1	07			'	7	G	W	g	w			§	·	▨	▨	ח	ק	7
1	0	0	0	08			(	8	H	X	h	x			¨	¸	▨	▨	ט	ר	8
1	0	0	1	09			)	9	I	Y	i	y			©	¹	▨	▨	י	ש	9
1	0	1	0	10			*	:	J	Z	j	z			×	÷	▨	▨	ך	ת	A
1	0	1	1	11			+	;	K	[	k	{			«	»	▨	▨	כ	▨	B
1	1	0	0	12			,	<	L	\\	l	|			¬	¼	▨	▨	ל	▨	C
1	1	0	1	13			–	=	M	]	m	}			SHY	½	▨	▨	ם	LRM	D
1	1	1	0	14			.	>	N	^	n	~			®	¾	▨	▨	מ	RLM	E
1	1	1	1	15			/	?	O	_	o				‾	▨	▨	=	ן	▨	F
					0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	hex

99-0097-A

7 Identification of the character set

7.1 Identification according to ECMA-35 and ECMA-43

The graphic characters of this ECMA Standard constitute a single coded character set. However, in accordance with ECMA-35 and ECMA-43 the code table of this ECMA Standard may 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 the identification methods of ECMA-35 or ECMA-43 are used, this ECMA Standard shall be identified by the following pair of designation functions:

GZD4   04/02   (ESC 02/08 04/02)

G1D6   04/07   (ESC 02/13 05/14)

NOTE

The corresponding escape sequences are shown in parentheses.

7.2   Identification using the ISO International register of coded character sets to be used with escape sequences

According to 7.1 above the character set of this ECMA Standard may be considered to consist of the character SPACE, a 94-character G0 graphic character set, and a 96-character G1 graphic character set. The G0 and G1 graphic character sets may be identified by the use of the Registration Numbers from the ISO International register of coded character sets to be used with escape sequences.

When these registration numbers are used this ECMA Standard shall be identified by the following pair of registration numbers:

– G0 graphic character set ISO-IR 6

– G1 graphic character set ISO/IR 198

Annex A
(informative)

Coverage of languages

A.1 Languages of European origin written in Latin script

The following ECMA Standards specify coded character sets which comprise various different selections of characters based on the Latin alphabet. These sets are identified by the numbers 1 to 6 as shown:

ECMA-94   Latin alphabets No. 1 to 4
ECMA-128   Latin alphabet No. 5
ECMA-144   Latin alphabet No. 6

The following official and regional languages written in Europe are covered by the Latin alphabets 1 to 6 as indicated by their number in table A.1:

Table A.1 - Language coverage

Language	Covered by alphabet(s)	Language	Covered by alphabet(s)	Language	Covered by alphabet(s)
Albania	1 2 5	Frisian	1 5	Norwegian	1 4 5 6
Basque	1 5	Galician	1 5	Polish	2
Breton	1 5	German	1 2 3 4 5 6	Portuguese	1 3 5
Catalan	1 5	Greenlandic	1 4 5 6	Rhaeto-Romanic	1 5
Croat	2	Hungarian	2	Romanian	2
Czech	2	Icelandic	1 6	Sámi	4 6
Danish	1 4 5 6	Irish Gaelic	1 5 6	Scottish Gaelic	1 5
Dutch	1 5	*(new orthography)*		Slovak	2
English	1 2 3 4 5 6	Italian	1 3 5	Slovene	2 4 6
Esperanto	3	Latin	1 2 3 4 5 6	Serbian	2
Estonian	4 6	Latvian	4	Spanish	1 5
Faroese	1 6	Lithuanian	4 6	Swedish	1 4 5 6
Finnish	1 4 5 6	Luxemburgish	1 5	Turkish	(3) 5
French	(1) (3) (5)	Maltese	3		

NOTES

1. The list of languages in table A.1 is not exhaustive. It shows the languages that are included in the Scope clause of the Latin alphabets.

2. For writing French, three characters (Œ, œ, Ÿ) not specified in Latin alphabets 1, 3 and 5, are also needed.

3. The various Sámi languages use partly differing orthographies. The character sets in Latin alphabets No. 4 and No. 6 cover the requirements of the Sámi languages most commonly used in Finland, Norway and Sweden. For the Skolt Sámi language used in Finland and Norway additional characters are needed.

4. There are several official written languages outside Europe that are covered by Latin alphabet No. 1. Examples are Indonesian/Malay, Tagalog (Philippines), Swahili, Afrikaans.

5. Use of Latin alphabet No. 3 for Turkish is deprecated.

A.2   Languages written in non-Latin scripts

The following standards specify coded character sets which include graphic characters from alphabets other than the Latin alphabet:

    ECMA-113   Latin/Cyrillic alphabet
    ECMA-114   Latin/Arabic alphabet
    ECMA-118   Latin/Greek alphabet
    ECMA-121   Latin/Hebrew alphabet

The following official and regional languages are covered by these alphabets:

Cyrillic characters included in Standard ECMA-113 cover Bulgarian, Byelorussian, (Slavic) Macedonian, Russian, Serbian and Ukranian (as written up to 1990, see also the Scope of Standard ECMA-113).

The Arabic characters included in .Standard ECMA-114 cover Arabic. The Greek characters included in ECMA-118 cover Greek (*monotonikó* orthography). The Hebrew characters included in ECMA-121 cover Hebrew.

Annex B
(informative)

Main differences between the first edition and this second edition of ECMA-121

B.1 The names of the graphic characters have been amended where necessary to align them with the names of the characters adopted for all standards on coded character sets developed under the responsibility of ISO/IEC JTC 1. For each character the short identifiers specified in ISO/IEC 10646-1, Amendment 9, have been added to table 1.

B.2 The new style of conformance clause, adopted for all standards on coded character sets, has been introduced.

B.3 Object identifiers conforming to Abstract Syntax Notation One (ASN.1, see ISO/IEC 8824-1) are specified in annex E for the character set, and the corresponding coded representations of this ECMA Standard.

Registration numbers from the International register of coded character sets to be used with escape sequences have been included as an additional method of identifying the coded character set of this ECMA Standard.

B.4 A new annex A has been added that identifies the coverage of languages by all Latin alphabets.

B.5 Various editorial adjustments and clarifications have been made to the text of the Standard. The hexadecimal equivalents of the bit combinations have been added to tables 1 and 2.

B.6 Support for bi-directionality has been included and is described in annex C.

B.7 Annex D, Bibliography, has been added.

Annex C
(informative)

Bi-directional text support

C.1   Bi-directional text formatting

The LEFT-TO-RIGHT MARK and RIGHT-TO-LEFT MARK characters are used in formatting bi-directional text. Text conforming to this ECMA Standard is typically rendered with an implicit bi-directional algorithm.

An implicit algorithm uses the directional character properties to determine the correct display order of characters on a horizontal line of text.

The characters specified in 5.3.4 and 5.3.5 are acting exactly like left-to-right or right-to-left characters in terms of affecting ordering (bi-directional format marks). They have no visible graphic symbols, and they do not have any other semantic effect.

An algorithm supporting bi-directional text formatting is described in the Unicode standard.

This ECMA Standard does not preclude the use of other means to manage the text format, such as the use of control functions from ECMA-48, or other means external to the graphic character set.

C.2   Backward compatibility

The two characters LEFT-TO-RIGHT MARK and RIGHT-TO-LEFT MARK are used in applications that have a bi-directional capability, i.e. applications supporting implicit directionality.

The behaviour of applications not supporting implicit directionality upon receiving of these characters is outside the scope of this ECMA Standard and therefore undefined.

Annex D
(informative)

Bibliography

ECMA-48   Control Functions for Coded Character Sets (1991)

ECMA TR/53   Handling of bi-directional texts (1992)

ISO/IEC 10646-1:1993 - Information technology -Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane

ISO International register of coded character sets to be used with escape sequences

The Standards Institution of Israel: SI 1311 (July 1989 - in revision), Information technology - ISO 8-bit coded character set for information interchange

The Unicode Consortium, The Unicode Standard - Version 2.0 (1996).