INTERNATIONAL STANDARD

ISO/IEC 2022

Fourth edition
1994-12-01

Information technology — Character code structure and extension techniques

Technologies de l'information — Structure de code de caractères et techniques d'extension

Reference number
ISO/IEC 2022:1994(E)

Contents

Foreword

ISO (the International Organisation for Standardisation) and IEC (the International Electrical Commission) form the specialised system for world-wide standardisation. National Bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organisation to deal with particular fields of mutual interest. Other international organisations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75% of the national bodies casting a vote.

International Standard ISO/IEC 2022 was prepared by the European Association for the Standardization of Information and Communication Systems, ECMA, (as ECMA-35) and was adopted, under a special "fast-track procedure", by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, in parallel with its approval by national bodies of ISO and IEC.

This fourth edition cancels and replaces the third edition (ISO 2022:1986), of which it constitutes a technical revision (see also the introduction).

Annex A forms an integral part of this International Standard. Annexes B, C and D are for information only.

Introduction

ECMA/TC1 participates very actively in the work of JTC1/SC2 (previously ISO/TC97/SC2) on code structure and code extension, and contributed numerous technical papers to SC2/WG1, the group entrusted with the preparation of ISO 2022, the International Standard for code extension techniques. ECMA published its first Standard ECMA-35 on the same subject in 1971. Three further editions in 1980, 1982 and 1985 reflected the progress achieved internationally, and the text of the 1985 edition was identical with that of the 1986 edition of ISO 2022.

The present edition of ISO/IEC 2022 is technically almost identical with the 1986 edition but is completely rearranged and rewritten to make it more convenient to use as a reference document.

This page intentionally left blank

Information technology - Character code structure and extension techniques

Section 1 - General

1 Scope

This International Standard specifies the structure of 8-bit codes and 7-bit codes which provide for the coding of character sets. The code elements used in the structure are common to both the 8-bit and 7-bit codes. The codes use a variety of techniques for extending the capabilities of elementary 8-bit and 7-bit codes. Greater emphasis is given to 8-bit codes in this edition of the Standard than in previous editions because they are now more widely used.

The use of common elements in the 8-bit and 7-bit code structure enables any specific conforming 8-bit code to be transformed into an equivalent 7-bit code, and vice versa, in a simple and direct fashion.

ISO/IEC 4873 conforms to the 8-bit code structure specified here, and ISO/IEC 646 conforms to the 7-bit code structure specified here.

Note - The coded character set specified in ISO/IEC 10646-1 has a different structure not in accordance with this International Standard.

The code structure facilities specified here include various means of extending the number of control functions and graphic characters available in a code. They also include techniques to construct and formalize the definition of specific codes, and to provide a coded identification of the structure and of the constituent elements of such specific codes.

Specific codes may also be identified by means of object identifiers in accordance with ISO 8824, Abstract Syntax Notation One (ASN.1). The form of such object identifiers is specified in annex A.

Individual character sets and control functions intended for use with these 8-bit and 7-bit codes are assumed to be registered in the ISO International Register of Coded Character Sets to be Used with Escape Sequences, in accordance with ISO 2375 (see annex B). The register includes details to relate individual character sets and control functions with their coded representations, and also with the associated coded identifications of such character sets.

The principles established in this International Standard may be utilized to form supplementary code structure facilities. For example ISO/IEC 6429 has followed such a procedure to formulate some parameterized control functions.

The use of uniform code structure techniques for the 8-bit and 7-bit codes specified here has the advantage of:

- permitting uniform provision for code structure in the design of information processing systems,
- providing standardized methods of calling into use agreed sets of characters,
- allowing the interchange of data between environments that utilise 8-bit and 7-bit codes respectively,
- reducing the risk of conflict between systems required to inter-operate.

When two systems with different levels of implementation of code structure facilities are required to communicate with one another, they may do so using the code structure facilities that they have in common.

The codes specified here are designed to be used for data that is processed sequentially in a forward direction. Use of these codes in strings of data which are processed in some other way, or which are included in data formatted for fixed-length record processing, may have undesirable results or may require additional special treatment to ensure correct interpretation.

Note - Since the previous edition (1986) of this International Standard the text has been completely rearranged and rewritten to make the Standard more convenient to use as a reference document. It is now arranged in three main sections as follows:

1 General

2 Character Sets and Codes

3 Code Identification and Escape Sequences

2 Conformance

2.1 Types of conformance

Full conformance to a standard means that all of its requirements are met. Conformance will only have a unique meaning if the standard contains no options. If there are options within the standard they must be clearly identified, and any claim of conformance must include a statement that identifies those options that have been adopted.

This International Standard is of a different nature since it specifies a large number of facilities from which different selections may be made to suit individual applications. These selections are not identified in this International Standard, but must be identified at the time that a claim of conformance is made. Conformance to such an identified selection is known as limited conformance.

The selection of facilities from this International Standard that are to be used in a particular application will generally be included in a specification document, which states the adopted facilities and gives other details necessary to define fully one or more specific codes. Such a specification is said to be in accordance with this International Standard (see 10.1).

2.2 Conformance of information interchange

A CC-data-element within coded information for interchange is in conformance with this International Standard if the coded representations within that CC-data-element satisfy the following conditions:

a) they shall represent graphic characters, control functions, and code-identification functions in accordance with an identified selection of the facilities specified in this International Standard (i.e. a version of this Standard, see 10.1);

b) when the code extension techniques specified in this International Standard are used, they shall be implemented by the control functions and code-identification functions defined in this Standard with the meaning and coded representation specified in this Standard;

c) no coded representation that is either reserved for registration and not assigned, or reserved for future use, shall be used;

d) no registered escape sequence shall be used with a meaning different from that defined by the registration.

2.3 Conformance of devices

A device is in conformance with this International Standard if it conforms to the requirements of 2.3.1, and either or both of 2.3.2 and 2.3.3 below. Any claim of conformance shall identify the document which contains the description specified in 2.3.1.

2.3.1 Device description

A device that conforms to this International Standard shall be the subject of a description that

a) identifies either directly, or by reference to a specification that is in accordance with this International Standard, the selection of facilities from this Standard that it can utilize when originating or when receiving CC-data-elements;

b) identifies the means by which the user may supply the corresponding characters and functions, or may recognize them when they are made available to the user, as specified in 2.3.2 and 2.3.3 respectively.

2.3.2 Originating devices

An originating device shall be capable of transmitting within a CC-data-element the coded representations of graphic characters from one or more graphic character sets, and of an identified selection of control functions and code-identification functions conforming to this International Standard.

Such a device shall allow the user to supply, from an appropriate set, characters or other indications which will implicitly or explicitly determine the graphic characters, control functions, and code-identification functions whose coded representations are to be transmitted.

2.3.3 Receiving devices

A receiving device shall be capable of receiving within a CC-data-element and interpreting the coded representations of graphic characters from one or more graphic character sets, and an identified selection of control functions and code-identification functions conforming to this International Standard.

Such a device shall make available to the user, from an appropriate set, characters or other indications which are implicitly or explicitly determined by the graphic characters, control functions, and code-identification functions whose coded representations are received.

3 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards listed below. Members of IEC and ISO maintain registers of currently valid standards.

ISO 2375:1985, *Data processing - Procedure for registration of escape sequences.*

ISO/IEC 6429:1992, *Information technology - Control functions for coded character sets.*

ISO 8824:1990, *Information technology - Open Systems Interconnection - Specification of Abstract Syntax Notation One (ASN.1).*

ISO 8825:1990, *Information technology - Open Systems Interconnection - Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1).*

ISO International Register of Coded Character Sets to be Used with Escape Sequences.

4 Definitions

For the purposes of this International Standard, the following definitions apply.

4.1 bit combination: An ordered set of bits used for the representation of characters.

4.2 byte: A bit string that is operated upon as a unit.

Note - Each bit has the value either ZERO or ONE.

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

4.4 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.

Notes

1 - In a communication environment in accordance with the Reference Model for Open Systems Interconnection of ISO 7498, a CC-data-element will form all or part of the information that corresponds to the Presentation-Protocol-Data-Unit (PPDU) defined in that International Standard.

2 - When information interchange is accomplished by means of interchangeable media, a CC-data-element will form all or part of the information that corresponds to the user data, and not that recorded during formatting and initialization.

4.5 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.

4.6 code extension: The techniques for the encoding of characters that are not included in the character set of a given code.

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

4.8 combining character: A member of an identified subset of a coded character set, intended for combination with the preceding or following graphic character, or with a sequence of combining characters preceded or followed by a non-combining character.

4.9 control character: A control function the coded representation of which consists of a single bit combination.

4.10 control function: An action that affects the recording, processing, transmission or interpretation of data, and that has a coded representation consisting of one or more bit combinations.

4.11 to designate: To identify a set of characters that are to be represented, in some cases immediately and in others on the occurrence of a further control function, in a prescribed manner.

4.12 device: A component of information processing equipment which can transmit, and/or can receive, coded information within CC-data-elements.

Note - It may be an input/output device in the conventional sense, or a process such as an application program or a gateway function.

4.13 escape sequence: A string of bit combinations that is used for control purposes in code extension procedures. The first of these bit combinations represents the control function ESCAPE.

Note -In this International Standard ESCAPE is always referred to as a control character.

4.14 Final Byte: The bit combination that terminates an escape sequence or a control sequence.

4.15 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.

4.16 graphic symbol: A visual representation of a graphic character or of a control function.

4.17 Intermediate Byte: A bit combination which may occur between that of the control character ESCAPE and the Final Byte in an escape sequence.

4.18 to invoke: To cause a designated set of characters to be represented by the prescribed bit combinations whenever those bit combinations occur.

4.19 repertoire: A specified set of characters that are each represented by one or more bit combinations of a coded character set.

4.20 to represent:

a) To use a prescribed bit combination with the meaning of a character in a set of characters that has been designated and invoked; or

b) To use an escape sequence with the meaning of an additional control function.

4.21 user: A person or other entity that invokes the services provided by a device.

Notes

1 - This entity may be a process such as an application program if the "device" is a code convertor or a gateway function, for example.

2 - The characters, as supplied by the user or made available to the user, may be in the form of codes local to the device, or of non-conventional visible representations, provided that 2.3 above is satisfied.

5 Notation, code tables 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 bits of the bit combinations of the 7-bit code are identified by b_7, b_6, b_5, b_4, b_3, b_2 and b_1, where b_7 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 integers in binary notation, in the range 0 to 255 for the 8-bit code, and in the range 0 to 127 for the 7-bit code, 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

In this International Standard, the bit combinations are identified by notations of the form x/y, where x and y are numbers in the range 00 to 15.

The correspondence between the notations of the form x/y and the bit combinations consisting of the bits b_8 or b_7 to b_1 is as follows:

– x for the 8-bit code 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;

– x for the 7-bit code is the number represented by b_7, b_6, and b_5 where these bits are given the weights 4, 2 and 1 respectively;

– y 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 notations of the form x/y are the same as those used to identify code table positions, where x is the column number and y the row number (see 5.2).

5.2 Code tables

An 8-bit code table consists of 256 positions arranged in 16 columns and 16 rows. The columns and rows are numbered 00 to 15 (see figure 1).

A 7-bit code table consists of 128 positions arranged in 8 columns and 16 rows. The columns are numbered 00 to 07 and the rows 00 to 15 (see figure 1).

The code table positions are identified by notations of the form x/y, where x is the column number and y is the row number. By convention, leading zeroes are included in the column and row numbers (e.g. 02/01).

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 x/y, is the same as that of the corresponding bit combination.

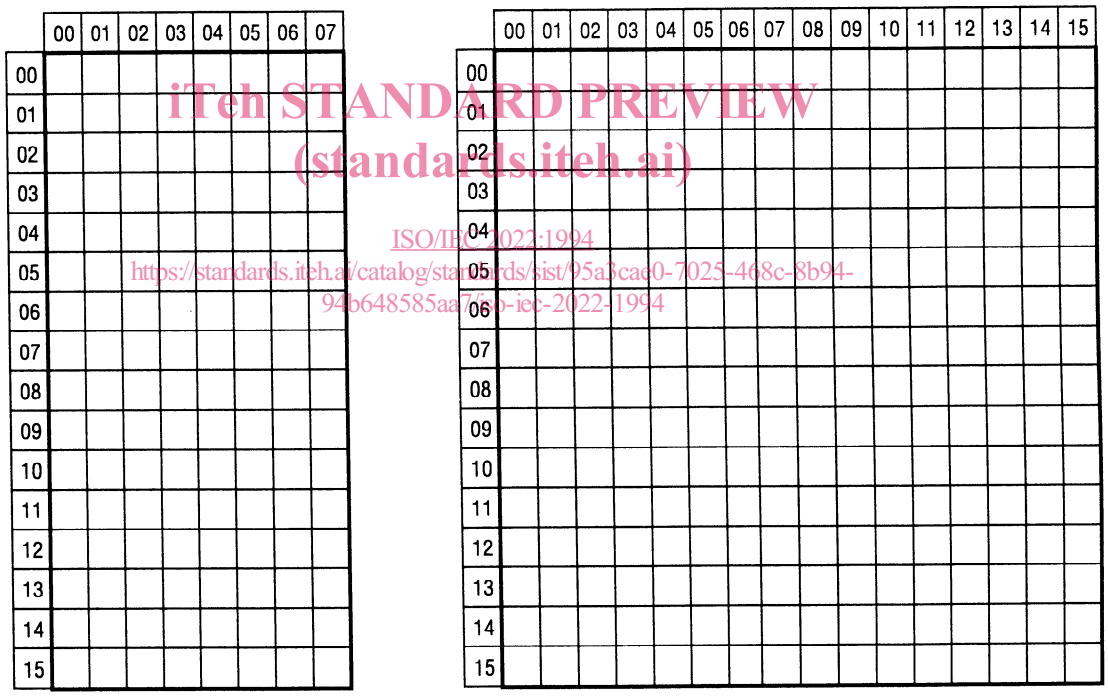

Figure 1 - Code tables

5.3 Names of characters

This International Standard assigns one name to each character. In addition, it specifies an acronym for each control character and for the characters SPACE and DELETE. By convention, only capital letters, space and hyphen are used for writing the names of the characters. For acronyms only capital letters and digits are used. It is intended that the acronyms and this convention be retained in all translations of the text.

Section 2 - Character sets and codes

6 Characters and character sets

6.1 Types of characters and character sets

The structure of 8-bit and 7-bit codes specified by this International Standard makes use of the following types of characters, character sets, and functions:

- fixed coded characters,
- sets of coded graphic characters,
- sets of coded control functions (or control characters),
- coded single additional control functions.

These components are specified respectively in 6.2 to 6.5 below.

The coded representations of the graphic characters and control functions are specified in relation to the 8-bit and 7-bit code tables defined in 5.2 above. A coded representation for each type of component is specified within columns 00 to 07 of the 8-bit and 7-bit code tables. For some components an alternative coded representation is specified in columns 08 to 15 of the 8-bit code table, and is not applicable to any 7-bit code.

6.2 Fixed coded characters

6.2.1 Character DELETE

Name: DELETE Acronym: DEL Coded representation: 07/15

DEL was originally used to erase or obliterate an erroneous or unwanted character in punched tape. DEL may be used for media-fill or time-fill. DEL characters may be inserted into, or removed from, a CC-data-element without affecting its information content, but such action may affect the information layout and/or the control of equipment.

6.2.2 Character ESCAPE

Name: ESCAPE Acronym: ESC Coded representation: 01/11

ESCAPE is a control character used for code extension purposes. It causes the meaning of a limited number of the bit combinations following it in a CC-data-element to be changed. These bit combinations, together with the preceding bit combination that represents the ESC character, constitute an escape sequence.

Escape sequences provide the coded representations of code-identification functions and of some types of control functions. The various uses of escape sequences are specified in clause 13. Code identification functions are specified in clauses 14 and 15.

6.2.3 Character SPACE

Name: SPACE Acronym: SP Coded representation: 02/00

SPACE is a graphic character. It has a visual representation consisting of the absence of a graphic symbol. It causes the active position to be advanced by one character position.

6.3 Sets of coded graphic characters

6.3.1 Types of coded graphic character set

A graphic character shall have a coded representation comprising one or more 8-bit combinations (bytes) in an 8-bit code, and one or more 7-bit combinations (bytes) in a 7-bit code. Within a coded graphic character set each character shall be represented by the same number of such bit combinations.

The bit combinations used to represent the graphic characters in a set shall be either from the six adjacent columns numbered 02 to 07 of the code tables or from the six adjacent columns numbered 10 to 15 of the 8-bit code table.

The type of a coded graphic character set is defined by the maximum number of graphic characters that the set can contain. The types of set specified here are illustrated in figure 3.

A coded graphic character set in which each character is represented by a single bit combination shall be one of the following:

- 94-character set, in positions 02/01 to 07/14, or 10/01 to 15/14;
  (i.e. all positions in columns 02 to 07 except 02/00 and 07/15, or
  all positions in columns 10 to 15 except 10/00 and 15/15)
- 96-character set, in positions 02/00 to 07/15, or 10/00 to 15/15.
  (i.e. all positions in columns 02 to 07, or in columns 10 to 15)

In a 94-character set no character shall be allocated to positions 02/00 and 07/15.

A coded graphic character set in which each character is represented by a sequence of n bit combinations, where n>1, shall be one of the following:

- 94^n-character set,
- 96^n-character set.

These sets are here referred to as multiple-byte sets.

A 94^n-character set shall consist of up to 94^n graphic characters each of which is represented by a sequence of n 8-bit or 7-bit combinations, either all in the range 02/01 to 07/14 or all in the range 10/01 to 15/14. In a 94^n-character set no character shall have a coded representation that includes the bit combination 02/00 or 07/15.

A 96^n-character set shall consist of up to 96^n graphic characters each of which is represented by a sequence of n 8-bit or 7-bit combinations, either all in the range 02/00 to 07/15 or all in the range 10/00 to 15/15.

Note - The 8th bit (b_8) of each byte in such an 8-bit multiple-byte representation is uniformly either ZERO or ONE.