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.

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.