INTERNATIONAL STANDARD

ISO/IEC 10646-1

First edition
1993-05-01

Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

Part 1:
Architecture and Basic Multilingual Plane

Technologies de l'information — Jeu universel de caractères codés à plusieurs octets —

Partie 1: Architecture et table multilingue

Reference number
ISO/IEC 10646-1:1993(E)

Contents

Annexes

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, 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 10646-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Sub-Committee SC 2, *Character sets and information coding*.

ISO/IEC 10646 consists of the following parts, under the general title *Information technology — Universal Multiple-Octet Coded Character Set (UCS)*:

— *Part 1: Architecture and Basic Multilingual Plane*

Additional parts will specify other planes.

Annexes A and B form an integral part of this part of ISO/IEC 10646. Annexes C to N are for information only.

Introduction

ISO/IEC 10646 specifies the Universal Multiple-Octet Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input and presentation of the written form of the languages (scripts) of the world as well as additional symbols.

This part of ISO/IEC 10646 specifies the overall architecture and the Basic Multilingual Plane (BMP) of the UCS.

Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

Part 1:

Architecture and Basic Multilingual Plane

1 Scope

ISO/IEC 10646 specifies the Universal Multiple-Octet Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input and presentation of the written form of the languages of the world as well as additional symbols.

This part of ISO/IEC 10646 specifies the overall architecture, and

- defines terms used in ISO/IEC 10646;

- describes the general structure of the coded character set;

- specifies the Basic Multilingual Plane (BMP) of the UCS, and defines a set of graphic characters used in scripts and the written form of languages on a world-wide scale;

- specifies the names for the graphic characters of the BMP, and the coded representations;

- specifies the four-octet (32-bit) canonical form of the UCS: UCS-4;

- specifies a two-octet (16-bit) BMP form of the UCS: UCS-2;

- specifies the coded representations for control functions;

- specifies the management of future additions to this coded character set.

The UCS is a coding system different from that specified in ISO 2022. The method to designate UCS from ISO 2022 is specified in 17.2.

2 Conformance

2.1 General

Whenever Private Use characters are used as specified in ISO/IEC 10646, the characters themselves shall not be covered by these conformance requirements.

2.2 Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with ISO/IEC 10646 if

a) all the coded representations of graphic characters within that CC-data-element conform to clauses 6 and 7, to an identified form chosen from clause 14, and to an identified implementation level chosen from clause 15;

b) all the graphic characters represented within that CC-data-element are taken from those within an identified subset (clause 13);

c) all the coded representations of control functions within that CC-data-element conform to clause 16.

A claim of conformance shall identify the adopted form, the adopted implementation level and the adopted subset by means of a list of collections and/or characters.

2.3 Conformance of devices

A device is in conformance with ISO/IEC 10646 if it conforms to the requirements of item a) below, and either or both of items b) , and c).

NOTE - The term *device* is defined (in 4.17) as a component of information processing equipment which can transmit and/or receive coded information within CC-data-elements. A device may be a conventional

input/output device, or a process such as an application program or gateway function.

A claim of conformance shall identify the document that contains the description specified in a) below, and shall identify the adopted form(s), the adopted implementation level, the adopted subset (by means of a list of collections and/or characters), and the selection of control functions adopted in accordance with clause 16.

a) Device description: A device that conforms to ISO/IEC 10646 shall be the subject of a description that identifies the means by which the user may supply characters to the device and/or may recognise them when they are made available to the user, as specified respectively, in subclauses b), and c) below.

b) Originating device: An originating device shall allow its user to supply any characters from an adopted subset, and be capable of transmitting their coded representations within a CC-data-element in accordance with the adopted form and implementation level.

c) Receiving device: A receiving device shall be capable of receiving and interpreting any coded representation of characters that are within a CC-data-element in accordance with the adopted form and implementation level, and shall make any corresponding characters from the adopted subset available to the user in such a way that the user can identify them.

Any corresponding characters that are not within the adopted subset shall be indicated to the user in a way which need not allow them to be distinguished from each other.

NOTES

1 An indication to the user may consist of making available the same character to represent all characters not in the adopted subset, or providing a distinctive audible or visible signal when appropriate to the type of user.

2 See also annex H for receiving devices with re-transmission capability.

3 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 10646. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO/IEC 10646 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 International Standards.

ISO 2022:1986 *Information processing — ISO 7-bit and 8-bit coded character sets —Code extension techniques.*

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