INTERNATIONAL STANDARD

ISO/IEC 10646-1

First edition
1993-05-01

AMENDMENT 9
1997-12-15

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

Part 1:
Architecture and Basic Multilingual Plane

AMENDMENT 9: Identifiers for characters

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

Partie 1: Architecture et table multilingue

AMENDEMENT 9: Identificateurs de caractères

Reference number
ISO/IEC 10646-1:1993/Amd.9:1997(E)

Documents
Document Sharing Hub

Full document details: ISO_IEC 10646-1_1993_AMD9_1997

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.

Amendment 9 to International Standard ISO/IEC 10646-1:1993 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

© ISO/IEC 1997

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

ISO/IEC Copyright Office • Case postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

Documents
Document Sharing Hub

Full document details: ISO_IEC 10646-1_1993_AMD9_1997

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

Part 1:

Architecture and Basic Multilingual Plane

AMENDMENT 9: Identifiers for characters

Page 4

Clause 6.2 Coding of characters

Replace final paragraph which reads:

When referring to characters within a plane, the leading four zeros (for G-octet and P-octet) may be omitted. For example, 0030 may be used to refer to DIGIT ZERO.

with:

When referring to characters within an identified plane, the leading four digits (for G-octet and P-octet) may be omitted. For example, within plane 00, 0030 may be used to refer to DIGIT ZERO.

Page 7

After clause **6.3 Octet order** *add two new clauses as follows.*

6.4 Naming of characters

ISO/IEC 10646 assigns a unique name to each character. The name of a character either:
a. denotes the customary meaning of the character, or
b. describes the shape of the corresponding graphic symbol, or
c. follows the rule given in clause 26 for Chinese/Japanese/Korean unified ideographs.

Guidelines to be used for constructing the names of characters in cases a. and b. are given in annex K.

6.5 Identifiers for characters

ISO/IEC 10646 defines a short identifier for each character. The short identifier for any character is distinct from the short identifier for any other character. These short identifiers are independent of the language in which this standard is written, and are thus retained in all translations of the text.

The following alternative forms of notation of a short identifier are defined here.

a. The eight-digit form of short identifier shall consist of the sequence of eight hexadecimal digits that represents the code position of the character (see 6.2).

b. The four-digit form of short identifier shall consist of the last four digits of the eight-digit form. It is not defined if the first four digits of the eight-digit form are not all zeroes; that is, for characters allocated outside the Basic Multilingual Plane.

c. The character "-" (HYPHEN-MINUS) may, as an option, precede the 8-digit form of short identifier.

d. The character "+" (PLUS SIGN) may, as an option, precede the 4-digit form of short identifier.

e. The prefix letter "U" (LATIN CAPITAL LETTER U) may, as an option, precede any of the four forms of short identifier defined in a. to d. above.

The CAPITAL letters A to F, and U that appear within identifiers may be replaced by the corresponding SMALL letters.

The full syntax of the notation of a short identifier, in Backus-Naur form, is:

{ U | u } [{+}xxxx | {-}xxxxxxxx]

where "x" represents one hexadecimal digit (0 to 9, A to F, or a to f), for example:

-hhhhhhhh +kkkk
Uhhhhhhhh U+kkkk

where hhhhhhhh indicates the eight-digit form and kkkk indicates the four-digit form.

NOTES

1. As an example the identifier for LATIN SMALL LETTER LONG S (see Table 3) may be notated in any of the following forms:

0000017F	-0000017F	U0000017F	U-0000017F
017F	+017F	U017F	U+017F

Any of the capital letters may be replaced by the corresponding small letter.

AVA Documents Document Sharing Hub

Full document details: ISO_IEC 10646-1_1993_AMD9_1997