INTERNATIONAL STANDARD

ISO/IEC 10646-1

First edition
1993-05-01

AMENDMENT 2
1996-10-15

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

Part 1:
Architecture and Basic Multilingual Plane

AMENDMENT 2: UCS Transformation Format 8 (UTF-8)

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

Partie 1: Architecture et table multilingue

AMENDEMENT 2: Format de transformation UCS 8 (UTF-8)

Reference number
ISO/IEC 10646-1:1993/Amd.2:1996(E)

AVA Documents
Document Sharing Hub

Full document details: ISO_IEC 10646-1_1993_Amd 2_1996

Contents

Page

Foreword iii

Introduction iv

2 Conformance 1

5 General structure of the UCS 1

Annexes

F The use of "signatures" to identify UCS 1

M External references to character repertoires 1

R UCS Transformation Format 8 (UTF-8) 2

R.1 Features of UTF-8 2

R.2 Specification of UTF-8 2

R.3 Notation 4

R.4 Mapping from UCS-4 form to UTF-8 form 4

R.5 Mapping from UTF-8 form to UCS-4 form 4

R.6 Identification of UTF-8 5

R.7 Incorrect sequences of octets: Interpretation by receiving devices 5

© ISO/IEC 1996

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

AVA Documents
Document Sharing Hub

Full document details: ISO_IEC 10646-1_1993_Amd 2_1996

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 2 to International Standard ISO/IEC 10646-1:1993 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*.

AVA Documents
Document Sharing Hub

Full document details: ISO_IEC 10646-1_1993_Amd 2_1996

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 amendment to ISO/IEC 10646 specifies an additional transformation format, UTF-8. In UTF-8 all the characters of the UCS have a coded representation which is suitable for use in communications and other environments where some octet values of the code are assumed to have a fixed definition according to ISO/IEC 4873.

Documents
Document Sharing Hub
Full document details: ISO_IEC 10646-1_1993_Amd 2_1996