

INTERNATIONAL STANDARD

ISO/IEC
10646-1

First edition
1993-05-01

AMENDMENT 8
1997-12-15

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

Part 1: Architecture and Basic Multilingual Plane

AMENDMENT 8

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

Partie 1: Architecture et table multilingue

AMENDEMENT 8



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 8 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

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

Part 1:

Architecture and Basic Multilingual Plane

AMENDMENT 8

Add the following new annex:

Annex T (informative)

Procedure for the unification and arrangement of CJK Ideographs

The graphic character collection CJK UNIFIED IDEOGRAPHS in ISO/IEC 10646-1:1993 contains 20,902 ideographs (see clause 26). They are derived from over 54,000 ideographs which are found in various different national and regional standards for coded character sets (the "source codes").

This Annex describes how the ideographs in this standard are derived from the source codes by applying a set of unification procedures. It also describes how the ideographs in this standard are arranged in the sequence of consecutive code positions to which they are assigned.

The source code standards are shown below in four groups according to their origins. The groups are identified as the G-, T-, J-, and K-sources.

G-source:	GB2312-80, GB12345-90, GB7589-87*, GB7590-87*, GB8565-88*, General Purpose Hanzi List for Modern Chinese Language*
T-source:	TCA-CNS 11643-1986/1st plane, TCA-CNS 11643-1986/2nd plane, TCA-CNS 11643-1986/14th plane*
J-source:	JIS X 0208-1990, JIS X 0212-1990
K-source:	KS C 5601-1989, KS C 5657-1991

(A " * " after the reference number of a standard indicates that some of the ideographs included in that standard are not introduced into the unified collection.)

For the purposes of ISO/IEC 10646-1 a unification process is applied to the ideographic characters taken from the codes in the source groups. In this process single ideographs from two or more of the source groups are associated together, and a single code position is assigned to them in this standard. The associations are made according to a set of procedures that are described below. Ideographs that are thus associated are described here as "unified".

T.1. Unification procedure

T.1.1 Scope of unification

Ideographs that are unrelated in historical derivation (non-cognate characters) have not been unified.

士, 土

Example:

NOTE - The difference of shape between the two ideographs in the above example is in the length of the lower horizontal line. This is considered an actual difference of shape. Furthermore these ideographs have different meanings. The meaning of the first is "Soldier" and of the second is "Soil or Earth".

An association between ideographs from different sources is made here if their shapes are sufficiently similar, according to the following system of classification.

T.1.2 Two level classification

A two-level system of classification is used to differentiate (a) between abstract shapes and (b) between actual shapes determined by particular typefaces. Variant forms of an ideograph, which can not be unified, are identified based on the difference between their abstract shapes.

T.1.3 Procedure

A unification procedure is used to determine whether two ideographs have the same abstract shape or different ones. The unification procedure has two stages, applied in the following order:

- a) Analysis of component structure;
- b) Analysis of component features;

T.1.3.1 Analysis of component structure

In the first stage of the procedure the component structure of each ideograph is examined. A component of an ideograph is a geometrical combination of primitive elements. Alternative ideographs can be configured from the same set of components. Components can be combined to create a new component with a more complicated structure. An ideograph, therefore, can be defined as a component tree, where the top node is the ideograph itself, and the bottom nodes are the primitive elements. This is shown in Figure 1.

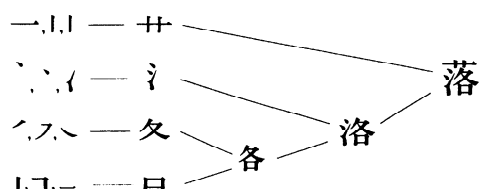


Figure 1 - Component structure

T.1.3.2 Analysis of component features

In the second stage of the procedure, the components located at corresponding nodes of two ideographs are compared, starting from the most superior node, as shown in Figure 2.

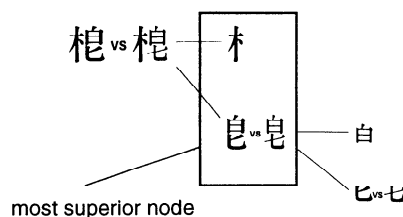


Figure 2 - The most superior node of a component

The following features of each ideograph to be compared are examined:

- a : the number of components,
- b : the relative position of the components in each complete ideograph,
- c : the structure of corresponding components.

If one or more of the features (a to c above) are different between the ideographs in the comparison, the ideographs are considered to have different abstract shapes and are therefore not unified.

If all of the features (a to c above) are the same between the ideographs, the ideographs are considered to have the same abstract shape and are therefore unified.

T.1.4 Examples of differences of abstract shapes

To illustrate rules derived from a: to c: in T.1.3.2, some typical examples of ideographs that are not unified, owing to differences of abstract shapes, are shown below.

T.1.4.1 Different number of components

The examples below illustrate rule a: since the two ideographs in each pair have different numbers of components.

崖·厓, 肱·肱, 降·夆

T.1.4.2 Different relative positions of components

The examples below illustrate rule b: Although the two ideographs in each pair have the same number of components, the relative positions of the components are different.

峰·峯, 荊·荊

T.1.4.3 Different structure of a corresponding component

The examples below illustrate rule c: The structure of one (or more) corresponding components within the two ideographs in each pair is different.