

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



Reference number
ISO/IEC 10646-1:1993/Amd.8:1997(E)

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.

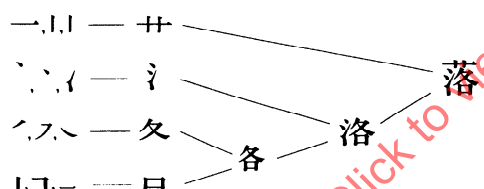


Figure 1 - Component structure

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.

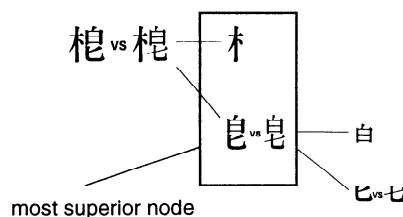


Figure 2 - The most superior node of a component

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.

扌·擴, 策·筭, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇

T.1.5 Differences of actual shapes

To illustrate the classification described in T.1.2, some typical examples of ideographs that are unified are shown below. The two or three ideographs in each group below have different actual shapes, but they are considered to have the same abstract shape, and are therefore unified.

𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇

The differences are further classified according to the following examples.

a) Differences in rotated strokes/dots

𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇

b) Differences in overshoot at the stroke initiation and/or termination

𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇

c) Differences in contact of strokes

𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇,
 𠂇·𠂇, 𠂇·𠂇

d) Differences in protrusion at the folded corner of strokes

𠂇·𠂇

e) Differences in bent strokes

𠂇·𠂇

f) Differences in folding back at the stroke termination

𠂇·𠂇

g) Differences in accent at the stroke initiation

𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇, 𠂇·𠂇

h) Differences in "rooftop" modification

𠂇·𠂇, 𠂇·𠂇

j) Combinations of the above differences

𠂇·𠂇

These differences in actual shapes of a unified ideograph are presented in the corresponding source columns for each code position entry in the code table in clause 26 of this International Standard.

T.1.6 Source separation rule

To preserve data integrity through multiple stages of code conversion (commonly known as "round-trip integrity"), any ideographs that are separately encoded in any one of the source standards listed above have not been unified.

However, some ideographs encoded in two standards belonging to the same source group (e.g. GB2312-80 and GB12345-90) have been unified during the process of collecting ideographs from the source group.

T.2. Arrangement procedure

T.2.1 Scope of arrangement

The arrangement of the CJK UNIFIED IDEOGRAPHS in the code table of clause 26 of this International Standard is based on the filing order of ideographs in the following dictionaries.

Priority	Dictionary	Edition
1	Kangxi Dictionary 康熙字典	Beijing 7th edition
2	Daikanwa Jiten 大漢和辭典	9th edition
3	Hanyu Dazidian 汉语大字典	1st edition
4	Daejaweon 大字源	1st edition

The dictionaries are used according to the priority order given in the table above. Priority 1 is highest. If an ideograph is found in one dictionary, the dictionaries of lower priority are not examined.

T.2.2 Procedure

T.2.2.1 Ideographs found in the dictionaries

a) If an ideograph is found in the Kangxi Dictionary, it is positioned in the code table in accordance with the Kangxi Dictionary order.

b) If an ideograph is not found in the Kangxi Dictionary but is found in the Daikanwa Jiten, it is given a position at the end of the radical-stroke group under which is indexed the nearest preceding Daikanwa Jiten character that also appears in the Kangxi dictionary.

c) If an ideograph is found in neither the Kangxi nor the Daikanwa, the Hanyu Dazidian and the Daejaweon dictionaries are referred to with a similar procedure.

T.2.2.2 Ideographs not found in the dictionaries

If an ideograph is not found in any of the four dictionaries, it is given a position at the end of the radical-stroke group (after the characters that are present in the dictionaries) and it is indexed under the same radical-stroke count.

T.3. Source code separation examples

The pairs (or triplets) of ideographs shown below are exceptions to the unification rules described in clause T.1 of this Annex. They are not unified because of the source code separation rule described in clause T.1.6.

NOTES

1. The particular source code group (or groups) that causes the source code separation rule to apply is indicated by the letter (G, J, K, or T) that appears to the right of each pair (or triplet) of ideographs. The source code groups that correspond to these letters are identified at the beginning of this Annex.

2. The ideograph pairs are listed below in ascending order by the code position of the first ideograph of each pair. The sequence progresses downwards in the left column as far as each marker (√-----√), and then continues downwards in the adjoining right column, starting at the previous marker.

丟丟	T	兗兗	T
4E1F 4E22		5156 5157	
么么	GT	冊冊	TJ
4E48 5E7A		518A 518C	
爭爭	GTJ	淨淨	G
4E89 722D		51C0 51C8	
仞仞	J	凡凡	T
4EDE 4EED		51E2 51E3	
併併	T	刃刃	TJ
4F75 5002		5203 5204	
侶侶	T	刊刊	TJ
4FA3 4FB6		520A 520B	
俣俣	TJK	刪刪	T
4FC1 4FE3		5220 522A	
兪兪	T	別別	T
4FDE 516A		5225 522B	
俱俱	T	券券	TJ
4FF1 5036		5238 52B5	
值值	T	剎剎	T
5024 503C		5239 524E	
偷偷	T	勗勗	T
5077 5078		524F 5259	
偽偽	TJ	剝剝	T
507D 50DE		525D 5265	
兌兌	T	劒劒	J
514C 5151		5292 5294	
兔兔	TJ	勻勻	T
514E 5154		52FB 5300	

√-----√

单单	T	墜墜	T	圈圈	TJ	媼媼	T
5355 5358		5848 588D		5708 570F		5A7E 5AAE	
即即	TK	填填	TJ	圓圓	T	媼媼	TK
5373 537D		5861 586B		570E 5713		5AAA 5ABC	
卷卷	TJ	增增	T	圖圖	T	媼媼	T
5377 5DFB		5897 589E		5716 5717		5AAF 5B00	
叁叁	GT	壯壯	GTJ	垚垚	T	媼媼	T
53C1 53C2		58EE 58EF		5759 5DE0		5B0E 5B14	
參參	T	壽壽	T	埤埤	J	嬾嬾	GT
53C3 53C4		58FD 5900		57D2 57D3		5B24 5B37	
呂呂	T	夂夂	T				
5415 5442		5910 657B		孳孳	T	弹弹	T
吞吞	T	本本	GTJ	5B73 5B76		5F39 5F3E	
541E 5451		5932 672C		宮宮	T	𠂇𠂇	TJ
吳吳吳	TJ	奧奧	J	5BAB 5BAE		5F50 5F51	
5433 5434 5449		5965 5967		寬寬	T	录录	T
吶吶	T	獎獎獎	TJ	5BDB 5BEC		5F54 5F55	
5436 5450		5968 596C 734E		寧寧	T	彙彙	T
告告	T	妝妝	GT	5BDC 5BE7		5F59 5F5A	
543F 544A		5986 599D		寢寢	GTJ	彝彝	J
唧唧	T	妍妍	T	5BDD 5BE2		5F5B 5F5C	
5527 559E		598D 59F8		專專	J	彝彝	T
噏噏	T	姍姍	T	5C02 5C08		5F5D 5F5E	
55A9 55BB		59CD 59D7		將將	GTJ	彥彥	T
噓噓	T	姬姬	GT	5C06 5C07		5F65 5F66	
5618 5653		59EB 59EC		尔尔	T	德德	T
噯噯	GTJ	娛娛娛	T	5C13 5C14		5FB3 5FB7	
568F 5694		5A1B 5A2F 5A31		尚尚	T	徵徵	T
囯囯	T	婕婕	T	5C19 5C1A		5FB4 5FB5	
56EF 56FD		5A55 5AAB		尕尕	T	惠惠	TJ
				5C2A 5C2B		6075 60E0	

檻檻	T	悅悅	T	揭揭	T	樣樣	TJ
5C36 5C37		6085 60A6		63B2 63ED		69D8 6A23	
屏屏	T	悞悞	T	搖搖搖	TJ	橫橫	T
5C4F 5C5B		609E 60AE		63FA 6416 6447		6A2A 6A6B	
崢崢	GT	憇憇	T	搵搵	T	步步	T
5CE5 5D22		60B3 60EA		63FE 6435		6B65 6B69	
巔巔	T	慍慍	T	擊擊	TJ	歲歲	T
5DD3 5DD4		6120 614D		6483 64CA		6B72 6B73	
餅餅	T	愼愼	TJ	教教	T	歿歿	T
5E21 5E32		613C 614E		654E 6559		6B7F 6B81	
帶帶	TJ	戩戩	GT	斂斂	T	殼殼	GTJ
5E2F 5E36		6229 622C		6553 655A		6BBB 6BBC	
并并	T	戲戲	T	既既	T	毀毀	T
5E76 5E77		622F 6231		65E2 65E3		6BC0 6BC1	
廢廢	T	戶戶戶	T	昂昂	T	每每	T
5EC4 5ECF		6236 6237 6238		6602 663B		6BCE 6BCF	
弑弑	T	戾戾	T	晚晚	T	氫氫	T
5F11 5F12		623B 623E		665A 6669		6C32 6C33	
強強	T	拋拋	T	暨暨	T	汚汚	T
5F37 5F3A		629B 62CB		66A8 66C1		6C5A 6C61	
v-----v							
拔拔	TJ	榆榆	T	曾曾	J	沒沒	TJ
629C 62D4		6961 6986		66FD 66FE		6C92 6CA1	
掙掙	T	概概	T	枊枊	T	淨淨	TJ
6329 635D		6982 69EA		67B4 67FA		6D44 6DE8	
插插插	TJ	楹楹	T	查查	T	涉涉	T
633F 63D2 63F7		6985 69B2		67E5 67FB		6D89 6E09	
捏捏	TJ	檣檣	T	柵柵	T	浼浼	T
634F 63D1		699D 6A27		67F5 6805		6D97 6D9A	
搜搜	TJ	槿槿	J	稅稅	T	淚淚	T
635C 641C		69C7 69D9		68B2 68C1		6D99 6DDA	
				v-----v			

淥淥	T	眾衆	TJK	瓶瓶	T	緼緼	T
6DE5 6E0C		773E 8846		74F6 7501		7E48 7E66	
清清	T	研研	T	產產	T	羹羹	TJ
6DF8 6E05		7814 784F		7522 7523		7FAE 7FB9	
渴渴	T	祿祿	TJ	瘦瘦	J	翹翹	T
6E07 6E34		797F 7984		75E9 7626		7FF6 7FFA	
溫溫	T	禿禿	T	皤皤	T	胼胼	T
6E29 6EAB		79BF 79C3		76A1 76A5		80FC 8141	
漚漚	T	稅稅	T	眞眞	TJ	脫脫	T
6E88 6F59		7A05 7A0E		771E 771F		812B 8131	
漑漑	T	穗穗	TJ	V-----V			
6E89 6F11		7A42 7A57		脛脛	T	謠謠	J
滾滾	T	箏箏	GJ	817D 8183		8B20 8B21	
6EDA 6EFE		7B5D 7B8F		烏烏	GT	𦍋𦍋	T
潛潛	GTJK	箴箴	T	8203 8204		8C5C 8C63	
6F5B 6FF3		7BB3 7C08		舍舍	TJ	走𠂔	TJ
瀨瀨	T	篡篡	T	820D 820E		8D70 8D71	
7028 702C		7BE1 7C12		舖舖	J	𨋖𨋖	T
為爲	GTJ	粵粵	T	8216 8217		8EFF 8F27	
70BA 7232		7CA4 7CB5		莊莊	TJ	輜輜	J
堯堯	GTJK	絕絕	T	8358 838A		8F1C 8F3A	
712D 7162		7D55 7D76		菑菑	TJ	輶輶	T
熙熙	J	綠綠	T	83D1 8458		8F3C 8F40	
7155 7199		7DA0 7DD1		莛莛	T	込込	T
煇煇	T	緒緒	T	8480 8495		8FBE 8FD6	
7174 7185		7DD2 7DD6		蔣蔣	GJ	迸迸	TJ
狀狀	GT	緣緣	T	848B 8523		8FF8 902C	
72B6 72C0		7DE3 7E01		蔦蔦	T	遙遙	J
瑤瑤	TJ	緼緼	T	848D 853F		9059 9065	
7464 7476		7DFC 7E15		蒨蒨	T	邢邢	T
				8570 8580		90A2 90C9	