

Universal Multiple-Octet Coded Character Set UCS
---

ISO/IEC JTC 1/SC 2/WG 2/P 5968/IRG N 1648 Confirmed

Date: 2010-03-16

Source:	IRG
Meeting:	After IRG#33, Danang, Vietnam
Title:	IRG Standing Document Summary
Status:	Member's submission
Actions required:	Feedback Requested
Distribution:	IRG
Medium:	Electronic
Pages:	5

**Introduction:**

IRG Working Document Series(IWDS) is a set of IRG maintained documents which keeps the up-to-date example lists of CJK unification related example cases to supplement the published Annex S of ISO/IEC 10646 for IRG unification work. The maintenance of the IWDS should comply with the operational procedures established in Annex E of the *IRG Principles and Procedures*.

In addition to this document(labeled as SUM) which explains the IWDS with a summary of the examples, there are also 3 document series as listed below:

Series 1: List of UCV (Unifiable Component Variations) of Ideographs(labeled as UCV)

Series 2: List of Non-Unifiable Components of Ideograph and Overly-Unified Ideographs(Labeled as UCV)

Series 3: List of Possibly Mis-Unified Ideographs(Labeled as MUI).

**File format of IWDS:**

Each of IWDS file is named as IWDS\_XXX\_II where XXX is the name of the specific series and II refers to the IRG meeting where the list is confirmed. Thus, unification example lists in the series which are updated from time to time will be taking the following format:

IWDS\_UCV: UCV list (Series 1)

IWDS\_NUC: Non-Unifiable Components (Series 2)

### **Detailed Specification of the Standing Document Series**

This section explains the nature of each series as well as the format and the information contained in each series.

#### 1. List of Unifiable Component Variations (UCV)

The UCV list provides the list of component variations to be unifiable observed from existing UCS multi-column charts, or proposed and agreed among IRG members to be unifiable.

If two ideographs differ only in terms of the components in the UCV list, but satisfy the requirement for dis-unification according to dis-unification rules, such ideographs may be encoded differently. However, such cases are exceptional and should be exhaustively listed in this document under the related components to avoid confusion for consideration of other characters.

Unification is meant to be at the component level only. In other words, if the components themselves are also ideographs proper, this list does not imply that the corresponding ideographs proper are unifiable.

The following is the format for each entry in the file:

- a. No.: The serial number of the entry for reference that is unique throughout the standardization works.
- b. Criteria: List of actual glyphs,
- c. References: Excerpt from existing document (e.g. JIS X 0213 and HYDZD)
- d. Exceptions: The exhaustive list of dis-unification examples.
- e. Compatible/Duplicate/Examples: The example list of unified ideographs and compatibility ideographs, and notes if necessary

The following is an example of a typical entry in UCV list.

47	朮 朮	朮 朮 (JIS X 0213 - 29) 朮⑤ 朮⑤ 朮麻 (HYDCCD - 22)	06037 06038	朮 朮 朮 朮 朮 朮 朮 朮 朮 朮	121/235	朮 朮 朮 朮	朮 2F957 朮 2FA01 朮 26CEF
			0672E 0672F		79EB 0-9F78 2-3150 1-5125 1-705C 0-7988 2-1748 1-4905 1-8060		
						朮 295B6 朮 朮 朮 132/129 8481 E-3A77 4-4036 1-5868 E-2687 4-3222 1-5672	
						朮 朮 朮 朮 朮 096/053 6035 0-6270 1-4D47 1-3D5D 2-3226 1-5534 0-6680 1-4539 1-2961 2-1806 1-5320	
						朮 朮 朮 108/173 6CAD 0-6370 2-2756 2-3F67 0-6780 2-0754 2-3171	
						朮 朮 112/162 70A2 3-4F5C 3-2D5C 3-4760 3-1360	
						朮 朮 朮 125/073 7D49 E-3827 3-3824 1-654E E-2407 3-2404 1-6946	
						朮 朮 朮 138/057 8A39 3-7270 2-403E 2-603B 3-9280 2-3230 2-6427	

## 2. List of Non-Unifiable Components of Ideograph(NUC)

The NUC list provides the list of component variations which are not to be unified. This list should be kept as minimum as possible. Components that are not obviously unifiable will not be listed here. That is, it should only list those that are close in glyph shapes which cognitively can be confusing. In other words, this list should only contain the components which are (possibly inappropriately) unified by precedence during the IRG working processing, or components that are stated to be unifiable by some local national standards, but not in the UCS.

Furthermore, this list should not contain components which are either (1) KangXi radicals (such as 工 vs. 土) or (2) simplified vs. traditional components with no precedence of unification (such as 門 vs. 门).

The following is the format for each entry in the file:

- Components: List of non-unifiable glyphs that is unique throughout the standardization works
- Analysis: Reasons for dis-unification and each reason will be listed separately. Typical ones are already separated ones and ideographs which are encoded by one-side only.
- Examples: List of exhaustive possibly over-unified ideographs (if exists)



