

Universal Multiple-Octet Coded Character Set UCS
---

ISO/IEC JTC 1/SC 2/WG 2/IRGN2234 Confirmed  
Date: 2017-10-19

Source:	IRG
Meeting:	After IRG Meeting #49, San Jose
Title:	IRG Standing Document Summary Version 2.
References	IRGN2234drafts, IRGN1648
Status:	Member's submission
Actions required:	Feedback Requested
Distribution:	IRG
Medium:	Electronic
Pages:	6

**Introduction:**

IRG Working Document Series (IWDS) is a set of IRG maintained documents to keep up-to-date examples of CJK unification related cases to supplement the published Annex S of ISO/IEC 10646 for IRG unification work. IRG also decided in its meeting #48 that a list of submitters's Printed Character Normalization Guidelines are to be kept in IWDS for keeping track of the transformation rules used for handwritten character to printed character conversions.

The maintenance of the IRG Working Document Series should comply with the operational procedures established in Annex E of the *IRG Principles and Procedures*.

The Standing Document Series consists of the following documents:

- Series 1: Summary of unification rules and sample examples.
- Series 2: List of Unifiable Component Variations (UCV) of Ideographs.
- Series 3: List of Non-Unifiable Components of ideographs and overly-unified ideographs (NUC)
- Series 4: List of possibly Mis-Unified Ideographs (MUI).
- Series 5: List of documents, each describes one submitter's normalization guidelines(SNG) for conversion of one particular script to the printed form of ideographs.

**File format of IWDS:**

Each of the IWDS file is named as IWDS\_ *SSS* \_ *II* where *SSS* is the name of the

specific series (may be more than 3 letters for Series 5) and *II* refers to the IRG meeting where the list is confirmed. Thus, the IRG working document series has 5 threads of documents. The first 4 threads are specified as follows:

IWDS\_SUM: Summary document (Series 1) ... This document.  
IWDS\_UCV: List of UCV (Series 2)  
IWDS\_NUC: List of NUC (Series 3)  
IWDS\_MUI: List of MUI (Series 4).

In Series 5, the guidelines are submitter dependent as well as script dependent. Thus, the file names need to include a submitter ID followed by a script ID. For example, the ROK's current guidelines are for conversion of cursive Kai to printed Kai style. So the file name should be: IWDS\_SNG\_ROK\_Kai\_49.

### **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. Summary of IRG Standing Document Series (SUM)**

The summary document (this document) is a definitive document giving detailed specification for each of the data files including specification on the nature of the data, the data format, and the examples.

#### **2. List of Unifiable Component Variations (UCV)**

The UCV list provides a list of component variations which are unifiable, as observed from existing UCS multi-column charts, or proposed and agreed among IRG members.

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, these ideographs may be encoded differently. However, these 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 this entry for reference that is unique throughout the standardization work.
- 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 the UCV list.

47	朮 朮 (JIS X 0213 - 29) 朮⑤ 朮⑤ 朮麻 (HYDZD - 22)	06037 06038 恣 恣 0672E 0672F 朮 朮	121/235 <b>79EB</b> <table style="display: inline-table; vertical-align: middle;"> <tr> <td>朮</td> <td>朮</td> <td>朮</td> <td>朮</td> </tr> <tr> <td>0-8F78</td> <td>2-3150</td> <td>1-5125</td> <td>1-705C</td> </tr> <tr> <td>0-7988</td> <td>2-1748</td> <td>1-4905</td> <td>1-8060</td> </tr> </table> 𠄎 295B6	朮	朮	朮	朮	0-8F78	2-3150	1-5125	1-705C	0-7988	2-1748	1-4905	1-8060	朮 2FA97 𠄎 2FA01 𠄎 28CEF			
朮	朮	朮	朮																
0-8F78	2-3150	1-5125	1-705C																
0-7988	2-1748	1-4905	1-8060																
			132/129 <b>8481</b> <table style="display: inline-table; vertical-align: middle;"> <tr> <td>𠄎</td> <td>𠄎</td> <td>𠄎</td> </tr> <tr> <td>E-3A77</td> <td>4-4036</td> <td>1-5868</td> </tr> <tr> <td>E-2687</td> <td>4-3222</td> <td>1-5672</td> </tr> </table>	𠄎	𠄎	𠄎	E-3A77	4-4036	1-5868	E-2687	4-3222	1-5672							
𠄎	𠄎	𠄎																	
E-3A77	4-4036	1-5868																	
E-2687	4-3222	1-5672																	
			096/053 <b>6035</b> <table style="display: inline-table; vertical-align: middle;"> <tr> <td>𠄎</td> <td>𠄎</td> <td>𠄎</td> <td>𠄎</td> <td>𠄎</td> </tr> <tr> <td>0-8270</td> <td>1-4D47</td> <td>1-3D5D</td> <td>2-3226</td> <td>1-5534</td> </tr> <tr> <td>0-8680</td> <td>1-4539</td> <td>1-2961</td> <td>2-1806</td> <td>1-5320</td> </tr> </table>	𠄎	𠄎	𠄎	𠄎	𠄎	0-8270	1-4D47	1-3D5D	2-3226	1-5534	0-8680	1-4539	1-2961	2-1806	1-5320	
𠄎	𠄎	𠄎	𠄎	𠄎															
0-8270	1-4D47	1-3D5D	2-3226	1-5534															
0-8680	1-4539	1-2961	2-1806	1-5320															
			108/173 <b>6CAD</b> <table style="display: inline-table; vertical-align: middle;"> <tr> <td>𠄎</td> <td>𠄎</td> <td>𠄎</td> </tr> <tr> <td>0-6370</td> <td>2-2756</td> <td>2-3F67</td> </tr> <tr> <td>0-8780</td> <td>2-0754</td> <td>2-3171</td> </tr> </table>	𠄎	𠄎	𠄎	0-6370	2-2756	2-3F67	0-8780	2-0754	2-3171							
𠄎	𠄎	𠄎																	
0-6370	2-2756	2-3F67																	
0-8780	2-0754	2-3171																	
			112/162 <b>70A2</b> <table style="display: inline-table; vertical-align: middle;"> <tr> <td>𠄎</td> <td>𠄎</td> </tr> <tr> <td>3-4F5C</td> <td>3-2D5C</td> </tr> <tr> <td>3-4760</td> <td>3-1360</td> </tr> </table>	𠄎	𠄎	3-4F5C	3-2D5C	3-4760	3-1360										
𠄎	𠄎																		
3-4F5C	3-2D5C																		
3-4760	3-1360																		
			125/073 <b>7D49</b> <table style="display: inline-table; vertical-align: middle;"> <tr> <td>𠄎</td> <td>𠄎</td> <td>𠄎</td> </tr> <tr> <td>E-3827</td> <td>3-3824</td> <td>1-654E</td> </tr> <tr> <td>E-2407</td> <td>3-2404</td> <td>1-6946</td> </tr> </table>	𠄎	𠄎	𠄎	E-3827	3-3824	1-654E	E-2407	3-2404	1-6946							
𠄎	𠄎	𠄎																	
E-3827	3-3824	1-654E																	
E-2407	3-2404	1-6946																	
			138/057 <b>8A39</b> <table style="display: inline-table; vertical-align: middle;"> <tr> <td>𠄎</td> <td>𠄎</td> <td>𠄎</td> </tr> <tr> <td>3-7270</td> <td>2-403E</td> <td>2-603B</td> </tr> <tr> <td>3-8280</td> <td>2-3230</td> <td>2-6427</td> </tr> </table>	𠄎	𠄎	𠄎	3-7270	2-403E	2-603B	3-8280	2-3230	2-6427							
𠄎	𠄎	𠄎																	
3-7270	2-403E	2-603B																	
3-8280	2-3230	2-6427																	

### 3. List of Non-Unifiable Components (NUC) of Ideographs

The NUC list provides a list of component variations which are not to be unified. This list should be kept as minimal 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 shape and can be confusing cognitively. In other words, this list should only contain components which are (possibly inappropriately) unified by precedence during the IRG working process, and 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 (1) KangXi radicals (such as 亠 vs. 士) or (2) simplified vs. traditional components with no precedence of unification (such as 門 vs. 阩).

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

- a. Components: List of non-unifiable glyphs that is unique throughout the standardization work.





- searching.
- b. Variant glyph(s): Actual glyph shape of the component in the source script
  - c. Normalized glyph: The corresponding normalized component glyph.
  - d. Evidence: Examples of actual character glyphs with reference to their sources.
  - e. Comments: Any remark that may be helpful to IRG review.

The following is an example of some typical entries in a normalization table(from ROK: IRGN2154V1.1).

SN (Serial Number)	Variant component shape	Normalized component shape	Evidences, examples, source	comments
1-1	步(少)	步(少)	頻-頻 [A50:14d;5(甲)] 賓-賓 [A272:27b;8(丙)]	※ refer to UCV_N009.
1-2	少	少	砂-砂 [A272:27c;4(丙)] 陟-陟 [A104:175d;6(丙)]	
1-3	少	少	少-少 [A50:30b;7(甲)]	
1-4	少	少	少-少 [A40:445b;6(甲)]	
1-5	少	少	少-少 [B061:322d;8(乙)]	
1-6	少	少	少-少 [B094:516b;1(丙)]	
1-7	少	少	隲-隲 [B081:501a;4]	
1-8	尒(小)	尒(少)	歲-歲 [B069:097c;1(乙)]	※ refer to UCV_N015.

(End of document)