

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC 1/SC 2/WG 2/IRGN2426

Date: 2019-10-28

Source:	IRG
Meeting:	After IRG#53
Title:	IRG Standing Document Summary Version 3.
References	IRGN2234, IRGN1648(Only minor editorial
Status:	changes)
Actions required:	Member's submission
Distribution:	Feedback Requested
Medium:	IRG
Pages:	Electronic
	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 Component Variations of ideographs and overly-unified ideographs (NUCV)

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	<table border="1"> <tr> <td>朮</td> <td>朮</td> </tr> <tr> <td colspan="2">(JIS X 0213 - 29)</td> </tr> <tr> <td>朮⑤</td> <td>朮⑤ 休 麻</td> </tr> <tr> <td colspan="2">(HYDZD - 22)</td> </tr> </table>	朮	朮	(JIS X 0213 - 29)		朮⑤	朮⑤ 休 麻	(HYDZD - 22)		06037 06038 恣 恣 0672E 0672F 朮 朮	<table border="1"> <tr> <td>121/235</td> <td>秫</td> <td>秫</td> <td>秫</td> <td>秫</td> <td></td> </tr> <tr> <td>79EB</td> <td>0-8F78</td> <td>2-3150</td> <td>1-5125</td> <td>1-705C</td> <td></td> </tr> <tr> <td></td> <td>0-7988</td> <td>2-1748</td> <td>1-4905</td> <td>1-8060</td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>秫</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>2F957</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>颯</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>2FA01</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>逌</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>26CEF</td> </tr> <tr> <td>132/129</td> <td>逌</td> <td>逌</td> <td>逌</td> <td></td> <td></td> </tr> <tr> <td>8481</td> <td>E-3A77</td> <td>4-4036</td> <td>1-5868</td> <td></td> <td></td> </tr> <tr> <td></td> <td>E-2687</td> <td>4-3222</td> <td>1-5672</td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>逌</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>26CEF</td> </tr> <tr> <td>096/053</td> <td>怵</td> <td>怵</td> <td>怵</td> <td>怵</td> <td>怵</td> </tr> <tr> <td>6035</td> <td>0-8270</td> <td>1-4D47</td> <td>1-3D5D</td> <td>2-3226</td> <td>1-5534</td> </tr> <tr> <td></td> <td>0-8680</td> <td>1-4539</td> <td>1-2961</td> <td>2-1806</td> <td>1-5320</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>108/173</td> <td>洙</td> <td>洙</td> <td></td> <td>洙</td> <td></td> </tr> <tr> <td>6CAD</td> <td>0-8370</td> <td>2-2756</td> <td></td> <td>2-3F67</td> <td></td> </tr> <tr> <td></td> <td>0-8780</td> <td>2-0754</td> <td></td> <td>2-3171</td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>112/162</td> <td>烺</td> <td>烺</td> <td></td> <td></td> <td></td> </tr> <tr> <td>70A2</td> <td>3-4F5C</td> <td>3-2D5C</td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>3-4760</td> <td>3-1360</td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>125/073</td> <td>紉</td> <td>紉</td> <td></td> <td>紉</td> <td></td> </tr> <tr> <td>7D49</td> <td>E-3827</td> <td>3-3824</td> <td></td> <td>1-654E</td> <td></td> </tr> <tr> <td></td> <td>E-2407</td> <td>3-2404</td> <td></td> <td>1-6946</td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>138/057</td> <td>誄</td> <td>誄</td> <td></td> <td>誄</td> <td></td> </tr> <tr> <td>8A39</td> <td>3-7270</td> <td>2-403E</td> <td></td> <td>2-603B</td> <td></td> </tr> <tr> <td></td> <td>3-8280</td> <td>2-3230</td> <td></td> <td>2-6427</td> <td></td> </tr> </table>	121/235	秫	秫	秫	秫		79EB	0-8F78	2-3150	1-5125	1-705C			0-7988	2-1748	1-4905	1-8060							秫						2F957						颯						2FA01						逌						26CEF	132/129	逌	逌	逌			8481	E-3A77	4-4036	1-5868				E-2687	4-3222	1-5672								逌						26CEF	096/053	怵	怵	怵	怵	怵	6035	0-8270	1-4D47	1-3D5D	2-3226	1-5534		0-8680	1-4539	1-2961	2-1806	1-5320							108/173	洙	洙		洙		6CAD	0-8370	2-2756		2-3F67			0-8780	2-0754		2-3171								112/162	烺	烺				70A2	3-4F5C	3-2D5C					3-4760	3-1360										125/073	紉	紉		紉		7D49	E-3827	3-3824		1-654E			E-2407	3-2404		1-6946								138/057	誄	誄		誄		8A39	3-7270	2-403E		2-603B			3-8280	2-3230		2-6427	
朮	朮																																																																																																																																																																																																																	
(JIS X 0213 - 29)																																																																																																																																																																																																																		
朮⑤	朮⑤ 休 麻																																																																																																																																																																																																																	
(HYDZD - 22)																																																																																																																																																																																																																		
121/235	秫	秫	秫	秫																																																																																																																																																																																																														
79EB	0-8F78	2-3150	1-5125	1-705C																																																																																																																																																																																																														
	0-7988	2-1748	1-4905	1-8060																																																																																																																																																																																																														
					秫																																																																																																																																																																																																													
					2F957																																																																																																																																																																																																													
					颯																																																																																																																																																																																																													
					2FA01																																																																																																																																																																																																													
					逌																																																																																																																																																																																																													
					26CEF																																																																																																																																																																																																													
132/129	逌	逌	逌																																																																																																																																																																																																															
8481	E-3A77	4-4036	1-5868																																																																																																																																																																																																															
	E-2687	4-3222	1-5672																																																																																																																																																																																																															
					逌																																																																																																																																																																																																													
					26CEF																																																																																																																																																																																																													
096/053	怵	怵	怵	怵	怵																																																																																																																																																																																																													
6035	0-8270	1-4D47	1-3D5D	2-3226	1-5534																																																																																																																																																																																																													
	0-8680	1-4539	1-2961	2-1806	1-5320																																																																																																																																																																																																													
108/173	洙	洙		洙																																																																																																																																																																																																														
6CAD	0-8370	2-2756		2-3F67																																																																																																																																																																																																														
	0-8780	2-0754		2-3171																																																																																																																																																																																																														
112/162	烺	烺																																																																																																																																																																																																																
70A2	3-4F5C	3-2D5C																																																																																																																																																																																																																
	3-4760	3-1360																																																																																																																																																																																																																
125/073	紉	紉		紉																																																																																																																																																																																																														
7D49	E-3827	3-3824		1-654E																																																																																																																																																																																																														
	E-2407	3-2404		1-6946																																																																																																																																																																																																														
138/057	誄	誄		誄																																																																																																																																																																																																														
8A39	3-7270	2-403E		2-603B																																																																																																																																																																																																														
	3-8280	2-3230		2-6427																																																																																																																																																																																																														

3. List of Non-Unifiable Component Variants (NUCV) 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 冂 vs. 冂).

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

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

092/110
5C6E

中 中 中 中

0-6578 2-212B 0-5625 2-2E4C
0-6988 2-0111 0-5405 2-1444

𠂇 (U+2F878)

𠂇 7826 サ (玉籀)作可切
小ひだりの手。左(𠂇)に通ず。(説文)𠂇、左手也、象形。(段注)俗以左右爲ナ又字。
解 象形。戸を左手とし、しを右手とする。手のかたちにかたどる。一説に、反又(𠂇)に従ひ、指事といふ。後、𠂇の字を左の字とし、左(たすく)の義は依に作る。(説文通訓定聲)リ、按从ニ反又、指事、リ、所以左、又手、者也、經傳皆以、左爲之。
備考 中(𠂇)は別字。(康熙字典、辨似、二字相似)中音微、草木初生也。𠂇、同、左。

𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇

𠂇 7825 𠂇(集韻)敕列切
小曰め。めばえ。めばえる。めぐむ。(説文)中、艸木初生也、象一出形有枝莖也、讀若微。(六書故)中、草木之生也微、故从艸之半。𠂇(𠂇)𠂇(𠂇)の古字。(説文)中、古文或曰爲𠂇字。(集韻)𠂇、或作𠂇。
解 象形。くさのめばえた形にかたどる。一は伸びるさまをあらはし、兩旁は其の枝をあらはす。假借して艸の字とする。(説文)中、艸木初生也、象一出形有枝莖也、古文或曰爲𠂇字、段注)一、讀若、引而上行也、枝、謂兩旁、莖、枝柱、謂一也、過乎中、則爲、生、下垂、相則爲、𠂇、云云、凡云古文以爲某字者、此明二六書之假借。
備考 𠂇(𠂇)は別字。(康熙字典、辨似、二字相似)𠂇、音微、草木初生也。𠂇、同、左。

5. Set of Printed Character Normalization Guidelines

As the normalization guidelines are submitter/culture/language dependent as well as script dependent, each document should provide an overview of the scope of the guidelines, the major references and authoritative document sources from where the guidelines are derived. A set of rules can be described using text descriptions with ample examples for people in other language/culture environments to follow and to help with the review and acceptance of evidence. It can also serve as a possible unification/disunification guide for other submitters. Content of the normalization table should include, but not limited to, the following data for each entry:

- a. Serial numbers: A numbering system used internally for indexing and

- 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)