Introduction:

The IRG Working Documents(IWD) Series is a set of IRG maintained documents which keeps the up-to-date examples of CJK unification related example cases to supplement the published Annex S of ISO/IEC 10646 for IRG's unification work. This document describes the data format and their use for these working documents. The maintenance of the IRG Working Documents Series should comply with the operational procedures established in Appendix E of the *IRG Principles and Procedures*.

The Working Documents Series consists of the following documents:

Series 1. Summary of Unification Rules and Examples
Series 2. List of UCV (Unifiable Component Variations) of Ideographs
Series 3. List of Non-Unifiable Components of Ideographs and
            Overly-Unified Ideographs
Series 4. List of Possibly Mis-Unified Ideographs

Naming and File format of IWDS:

Each of IWD file is named as IRGWD_*SSS_II* where *SSS* is the name of the specific series and *II* refers to the IRG meeting where the list is confirmed. Thus, the Working Documents Series has 4 threads of documents as given below:

   IRGWD_SUM: Summary document (Series 1) … This document.
   IRGWD_UCV: UCV list (Series 2)
   IRGWD_NUC: Non-Unifiable Components (Series 3)

Detailed Specification of the Working 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 Working 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 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.
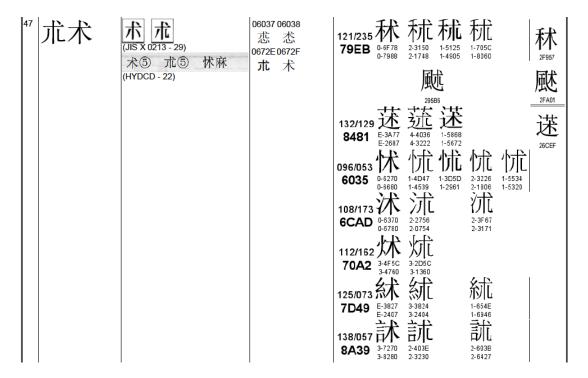
If two ideographs differ only in the components in the UCV list, but satisfy the requirement for dis-unification according to dis-unification rules, such ideographs may be encoded differently. However, such cases are exceptional and should be exhaustively listed in this document under the related components 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 such ideographs proper are unifiable.

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

  a. The serial number of this entry for reference, unique throughout the standardization process.

  b. List of actual glyphs.

  c. Excerpts from existing documents (e.g. JIS X 0213 and HYDZD).

  d. The exhaustive list of dis-unification examples.

  e. 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 术

术 术
(JIS X 0213 - 29)

术⑤　术⑤　怵麻
(HYDCD - 22)

06037 06038
怸　态
0672E 0672F
朮　术

121/235 秫 秫 秫 秫
79EB 0-6F78 2-3150 1-5125 1-705C
      0-7988 2-1748 1-4905 1-8060

颱
295B6

2F957 秫

2FA01 颱

132/129 莍 蒁 迷
8481 E-3A77 4-4036 1-5868
      E-2687 4-3222 1-5672

迷
26CEF

096/053 怵 怵 怵 怵 怵
6035 0-6270 1-4D47 1-3D5D 2-3226 1-5534
      0-6680 1-4539 1-2961 2-1806 1-5320

108/173 沭 沭 沭
6CAD 0-6370 2-2756 2-3F67
      0-6780 2-0754 2-3171

112/162 烌 烌
70A2 3-4F5C 3-2D5C
      3-4760 3-1360

125/073 絑 絉 絉
7D49 E-3827 3-3824 1-654E
      E-2407 3-2404 1-6946

138/057 �output �output �output
8A39 3-7270 2-403E 2-603B
      3-8280 2-3230 2-6427

3. 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 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 shapes which cognitively can be confusing. In other words, this list should only contain the components which are (possibly inappropriately) unified by precedent during the IRG working process, 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) a pair of KangXi radicals (such as 工 vs. 土) or (2) simplified vs. traditional components with no precedent of unification (such as 門 vs 门).

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

  a. List of non-unifiable glyphs.

  b. List of exhaustive possibly over-unified ideographs (if exists).

  c. List of sample ideographs that are already separately encoded.

  d. List of ideographs which are encoded by one-side only.

The following is an example of a typical entry in NUC:

Components	Analysis	Samples
厤 厯	Unifications	
	Separations	厤麻, 懇懇, 曆曆, 歷歷, 瀝瀝, 曆曆, 歷歷
	厤-only	none
	麻-only	曆厥歷厤厤厯厤厤厥摡歷曆厲厤曆歷曆曆厲曆魘鸝曆
閒 間	Unifications	喎 053/244 35F4 5-3856 5-2454 4-492D 4-4113 喎 喎 3-226F 3-0279 喎 喎 0-3338 0-1924 2-4-34 2442 a4c2 f1c0 35F4 喎 30 (口) 12 4277 ガン人名 マ 喎口(マグチ・姓) 喎 喎 喎
	Separations	個個, 嫻嫻, 憪憪, 捫捫, 楅楅, 潤潤, 爛爛, 癎癎, 瞷瞷, 磵磵, 簡簡, 繝繝, 萠萠, 襴襴, 覸覸, 調調, 調調, 鑭鑭, 閈閈, 簡簡, 鷴鷴, 嶰嶰, 鋼鋼, 蜎蜎, 醐醐, 鎖鎖, 鋼鋼, 鬮鬮
	間-only	喎堝喎珚楅喎蜎喎覸
	閒-only	胭殱猟鋼鋼鋼

4.   List of Possibly Mis-Unified Ideographs (MUI).

MUI list provides the possibly mis-unified ideographs as pairs of CJK compatibility ideographs and their corresponding CJK unified ideographs, which have different meanings and pronunciations with the supplied related reference information (possibly a dictionary) in a single document.

It is possible that the coded CJK compatibility ideographs listed in this document be proposed as new CJK unified ideographs. However, extreme care must be taken to assure the compatibility with existing standards in accordance with Annex I of WG2's Principles and Procedures.

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

  a. List of possibly non-unifiable ideographs.

  b. Excerpts of their usage from a single document source.

The following is an example of a typical entry in MUI:

中 中 屮 屮 屮 (U+2F878)

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

【屮】
7826

サ 〔玉篇〕作可切

ひだりの手。 左（4-8720）に通

〔說文〕屮、左手也、象形。〔段

注〕俗以二左右一爲二ナ又字一。

【解字】 象形。ﾉﾟを左手とし、ﾍﾟを右手とする。手の

かたちにかたどる。一說に、反又（ﾍﾟ）に從ひ、指

事といふ。後、屮の字を左の字とし、左（たすく）

の義は佐に作る。〔說文通訓定聲〕屮、按從二反

又一、指事、屮、所二以左又手一者也、經傳皆以レ左

爲レ之。

【參考】 中（4-7825）は別字。〔康熙字典、辨似〕二字

相似）中、音徹、草木初生也。屮、同レ左。

【屮】
7825

〔集韻〕敕列切 テツ チチ

〔集韻〕采早切 ゾウ サウ

め。めばえ。めぐ

〔說文〕屮、艸木初生也、象二

―出形有二枝莖一也、讀若レ徹。〔六書故〕

屮、草木之生也微、故从二艸之半一。〔―

30638〕の古字。〔說文〕屮、古文或目爲レ艸

字二。〔集韻〕艸、或作レ屮。

【解字】 象形。くさのめばえた形にかたどる。―は

伸びるさまをあらはし、兩旁は其の枝をあらは

す。假借して屮の字とする。―は〔說文、屮艸木初生

也象二―出一形有二枝莖一也古文或目爲二艸字一、段

注〕―、讀若レ凶、引而上行也、枝、謂二兩旁一、莖、枝

柱、謂レ一也、過二乎屮一則爲レ艸、下垂二相則爲レ枝、枝

木、云云、凡云二古文以爲二某字一者、此明二六書之

假借一。

【參考】 屮（4-7826）は別字。〔康熙字典、辨似〕二字

相似）中、音徹、草木初生也。屮、同レ左。