Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC 1/SC 2/WG 2 N 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 Ideograhs(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_*SSS_II* where *SSS* 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.
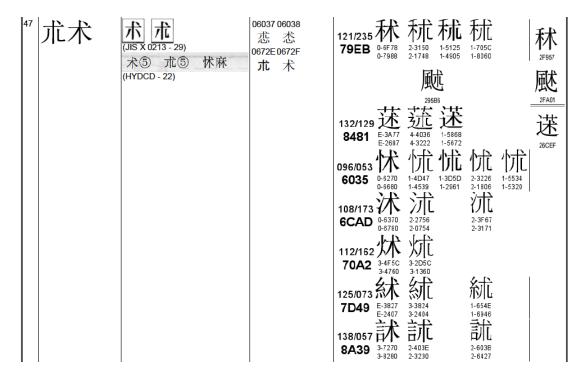
If two ideographs differ only in terms of the components in the UCV list, but satisfy the requirement for dis-unification according to dis-unification rules, such ideographs may be encoded differently. However, such cases are exceptional and should be exhaustively listed in this document under the related 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.

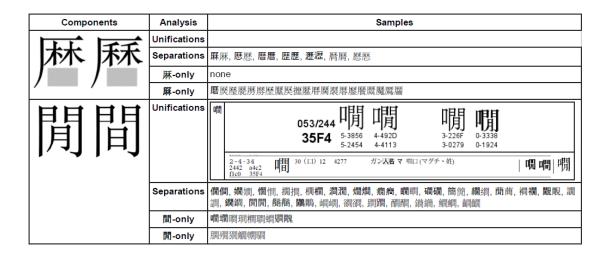## 2. List of Non-Unifiable Components of Ideograph(NUC)

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:

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

b. 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.

c. Examples: List of exhaustive possibly over-unified ideographs (if exists)

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

Components	Analysis	Samples
厤 厯	Unifications	
	Separations	厤厤, 厯厤, 厤厤, 歴歴, 歴歴, 厤厤, 歴歴
	厤-only	none
	厤-only	厤厤歴歴厤厤歴歴掫歴厤厤厤歴厤歴歴厤厤厤
閒 間	Unifications	
	Separations	個間, 嫺姻, 憪憪, 捫捫, 椚椚, 澗澗, 燗燗, 痫病, 瞷瞷, 硴硴, 簡簡, 緍緍, 萠萠, 襇襇, 覸覸, 調調, 鐧鐧, 閒閒, 鬜鬜, 鵳鵳, 蜠蜠, 礀礀, 睴睴, 醐醐, 鍆鍆, 綱綱, 鬪鬪
	間-only	喞塕閒珟椚睴蜠睴覸
	閒-only	胴殉猧鮦綢騳

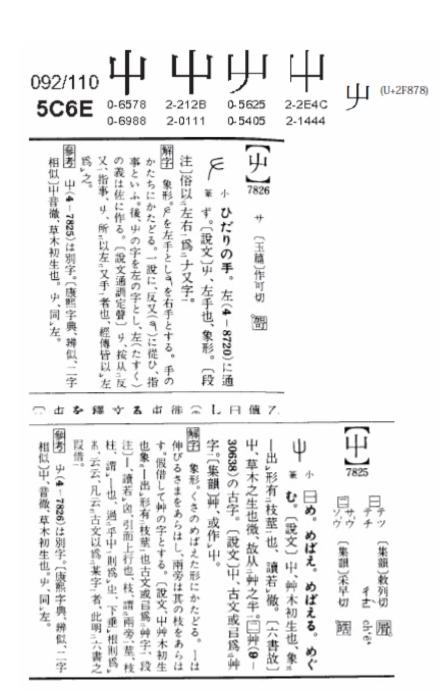3. List of Possibly Mis-Unified Ideographs(MUI).

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

It is possible that the coded CJK compatibility ideographs listed in this document need to 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. U-code: The UCS codepoint

  b. Characters: List of possibly non-unifiable ideographs.

  b. References: 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）に通ず。〔說文〕屮、左手也、象形。〔段注〕俗以左右二為二ナ又字。

[解字] 象形。ﾖを左手とし、ﾖ之を右手とする。手のかたちにかたどる。一說に、反又（ﾖ）を右手とし、左（たすく）に從ひ、指事。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、按从二反又一指事、屮、所以左二又手一者也、經傳皆以左為之。

[參考] 屮（4－7825）は別字。〔康熙字典、辨似二二字相似〕屮、音徹、草木初生也。屮、同二左。

【中】
7825

サウ

テツ
チ

〔集韻〕敕列切

〔集韻〕采早切

め。めばえ。めぐ。 〔說文〕中、艸木初生也、象二｜出形有二枝莖一也、讀若徹。〔六書故〕中、艸木之生也微、故从二艸之牛二30638〕の古字。〔說文〕中、古文或曰為二艸一字二〔集韻〕艸、或作中。

[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。假借して艸の字とする。｜は出二形有二枝莖一也古文或曰為二枝莖一、段

[注] 一、讀若、凶、引而上行也、枝、讀二兩旁一葉、枝柱、謂二一也、過二乎中一則為二屮、下垂二相則為木、云云、凡云二古文以為二某字一者、此明二六書之假借一。

[參考] 屮（4－7826）は別字。〔康熙字典、辨似二二字相似〕中、音徹、草木初生也。屮、同二左。

(End of document)