

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N [1107](#)

Date: 2005-01-20

Source:	Taichi Kawabata and Lu Qin
Title:	IDS Decomposition Principles
Status :	
Actions required	Review by IRG Editors for discussion at IRG meeting No. 24.
Distribution:	IRG Members and Ideographic Experts
Medium :	Electronic

Guidelines for composing IDS (Ideographic Description Sequence).

1. Introduction

IDS (Ideographic Description Sequence) is the character sequence which describes how the Ideograph is composed, by using the Ideographic Description Character (IDC) and Description Components (DC). For the usage of IDC and DC, please refer the Annex section F.3 and Table F.1 of ISO 10646:2003 (pp. 1736-1738) (it is also mentioned in IRG N955). In this document, we will describe the some detailed instruction on composing IDS.

2. Some Hints on Composition

2.1. Minimum decomposition principles

2.1.1. No-Radical-Decomposition Principle

For each ideograph proper, if it is listed in “[信息處理用 GB13000.1 字符集漢字部件規範](#)”(See attachment document), or if it is already a Kang Xi radical or a CJK Radical Supplement, it will not be further decomposed.

2.1.2. Shortest IDS Principle

Try to decompose a character using the shortest IDS possible. In other words, if the character's identified components are already coded ideographs, those component characters should not be further decomposed.

For example, 桂 → 𣎵木圭 (not 𣎵木𠩺土土)

森 → 𣎵木林 (not 𣎵木𣎵木木)

2.1.3. Nesting avoidance principle

Nest of IDS should be avoided as much as possible even though sometimes it cannot be avoided. (Note: If you try use the Shortest IDS Principle, minimum nesting would be achieved, so this rule may be redundant, but I have not time to prove it. So I am leaving it here)

2.2 Radical-Split Principle

If you wonder about the decomposition point, try splitting the character by the radicals first. After that, split the character to the visually separated components. Visually connected components should be splitted at last.

For example: 鴻 → 江鳥 (not 鴻 , nor 鴻)
(against minimum principle)
树 → 木对 (not 树 , nor 树)
(against minimum principle)

2.3. Choose-Binary-Symbol-First Principle

If the target character can be described by both the combination of binary IDC (such as 𠄎, 𠄎, 𠄎, etc) or ternary operator (𠄎 or 𠄎), we should choose the binary symbol as much as possible to achieve minimum IDS

For example, If it can follow Radical-split principle in 2.2, to split the character into two components, always use the binary operator.

Example : 壙 → 土恒 (not 土 亅 亅 (against minimum principle))

However, if the non-radical part is not a character, yet using the ternary operators can achieve minimum decomposition, use the ternary operator.

Example: 京 → 亠 日 小 (not 亠 亠 日 小)

弣 → 弓 百 弓 (not 弣 弓 as the radical should be on the left hand side, yet 百 弓 is not a coded ideograph; Also, 弓 百 弓 would not be minimum and further nesting would be needed)

鴻 → 灬九鳥 (not 灬 鴻 as it is against the Radical-split Principle; ; Also not , 灬 灬 灬 木九鳥 as it is against the Minimum-nesting Principle and the Minimal IDS Principle)

2.4 Overlap-Symbol-Avoidance Principle

As the Overlap Symbol (𠂇) is a somewhat fuzzy description, it should be avoided as much as possible. However, if a decomposition is unavoidable, try to respect the writing custom and single strokes should not be split into two to avoid using this symbol)

Example: 𠂇 → 𠂇 十从巾

Another example: 串 𠂇 口中 (not 中口, nor 𠂇 中中)

2.5. Surround-Over-Ternary-Symbol Principle.

If the character can be considered to have the surrounding radicals, use it as much as possible .

Example : 裏 → 𠂇 衣里 (not 亠 里 亠 (+UF7EE))

街 → 𠂇 行圭 (not 𠂇 彳 圭 亠)

but characters such as 章 should not be decomposed similarly into 𠂇 辛 日

It should only be used when two characters really overlaps. They shouldn't be used when the one surrounds other.

Example:

土 → 𠂇 土 丶 (not 𠂇 土 丶)

土 → 𠂇 土 丶 (not 𠂇 土 丶)

3. Prohibited Characters.

3.1 Kanxi Radical characters (U+2F00 - U+2FDF) should never be used. Instead, use the corresponding Ideograph characters.

一(U+2F00)	一(U+4E00)
丨(U+2F01)	丨(U+4E28)
丶(U+2F02)	丶(U+4E36)
丿(U+2F03)	丿(U+4E3F)
乙(U+2F04)	乙(U+4E59)
丿(U+2F05)	丿(U+4E85)
二(U+2F06)	二(U+4E8C)
亅(U+2F07)	亅(U+4EA0)
人(U+2F08)	人(U+4EBA)
儿(U+2F09)	儿(U+513F)
入(U+2F0A)	入(U+5165)
八(U+2F0B)	八(U+516B)
冂(U+2F0C)	冂(U+5182)
冂(U+2F0D)	冂(U+5196)
冂(U+2F0E)	冂(U+51AB)
几(U+2F0F)	几(U+51E0)
凵(U+2F10)	凵(U+51F5)
刀(U+2F11)	刀(U+5200)
力(U+2F12)	力(U+529B)
勹(U+2F13)	勹(U+52F9)
匕(U+2F14)	匕(U+5315)
匚(U+2F15)	匚(U+531A)
匚(U+2F16)	匚(U+5338)
十(U+2F17)	十(U+5341)
卜(U+2F18)	卜(U+535C)
冂(U+2F19)	冂(U+5369)
凵(U+2F1A)	凵(U+5382)
厶(U+2F1B)	厶(U+53B6)
又(U+2F1C)	又(U+53C8)
匚(U+2F1D)	匚(U+53E3)
匚(U+2F1E)	匚(U+56D7)
士(U+2F1F)	士(U+571F)
士(U+2F20)	士(U+58EB)
夕(U+2F21)	夕(U+5902)
夕(U+2F22)	夕(U+590A)
夕(U+2F23)	夕(U+5915)

大(U+2F24) 大(U+5927)
女(U+2F25) 女(U+5973)
子(U+2F26) 子(U+5B50)
宀(U+2F27) 宀(U+5B80)
寸(U+2F28) 寸(U+5BF8)
小(U+2F29) 小(U+5C0F)
尢(U+2F2A) 尢(U+5C22)
尸(U+2F2B) 尸(U+5C38)
巾(U+2F2C) 巾(U+5C6E)
山(U+2F2D) 山(U+5C71)
巛(U+2F2E) 巛(U+5DDB)
工(U+2F2F) 工(U+5DE5)
己(U+2F30) 己(U+5DF1)
巾(U+2F31) 巾(U+5DFE)
干(U+2F32) 干(U+5E72)
彡(U+2F33) 彡(U+5E7A)
广(U+2F34) 广(U+5E7F)
廴(U+2F35) 廴(U+5EF4)
井(U+2F36) 井(U+5EFE)
弋(U+2F37) 弋(U+5F0B)
弓(U+2F38) 弓(U+5F13)
厶(U+2F39) 厶(U+5F50)
彡(U+2F3A) 彡(U+5F61)
彳(U+2F3B) 彳(U+5F73)
心(U+2F3C) 心(U+5FC3)
戈(U+2F3D) 戈(U+6208)
戶(U+2F3E) 戶(U+6236)
手(U+2F3F) 手(U+624B)
支(U+2F40) 支(U+652F)
攴(U+2F41) 攴(U+6534)
文(U+2F42) 文(U+6587)
斗(U+2F43) 斗(U+6597)
斤(U+2F44) 斤(U+65A4)
方(U+2F45) 方(U+65B9)
无(U+2F46) 无(U+65E0)
日(U+2F47) 日(U+65E5)
日(U+2F48) 日(U+66F0)

月(U+2F49) 月(U+6708)
木(U+2F4A) 木(U+6728)
欠(U+2F4B) 欠(U+6B20)
止(U+2F4C) 止(U+6B62)
歹(U+2F4D) 歹(U+6B79)
殳(U+2F4E) 殳(U+6BB3)
毋(U+2F4F) 毋(U+6BCB)
比(U+2F50) 比(U+6BD4)
毛(U+2F51) 毛(U+6BDB)
氏(U+2F52) 氏(U+6C0F)
气(U+2F53) 气(U+6C14)
水(U+2F54) 水(U+6C34)
火(U+2F55) 火(U+706B)
爪(U+2F56) 爪(U+722A)
父(U+2F57) 父(U+7236)
爻(U+2F58) 爻(U+723B)
片(U+2F59) 片(U+723F)
片(U+2F5A) 片(U+7247)
牙(U+2F5B) 牙(U+7259)
牛(U+2F5C) 牛(U+725B)
犬(U+2F5D) 犬(U+72AC)
玄(U+2F5E) 玄(U+7384)
玉(U+2F5F) 玉(U+7389)
瓜(U+2F60) 瓜(U+74DC)
瓦(U+2F61) 瓦(U+74E6)
甘(U+2F62) 甘(U+7518)
生(U+2F63) 生(U+751F)
用(U+2F64) 用(U+7528)
田(U+2F65) 田(U+7530)
疋(U+2F66) 疋(U+758B)
疒(U+2F67) 疒(U+7592)
夨(U+2F68) 夨(U+7676)
白(U+2F69) 白(U+767D)
皮(U+2F6A) 皮(U+76AE)
皿(U+2F6B) 皿(U+76BF)
目(U+2F6C) 目(U+76EE)
矛(U+2F6D) 矛(U+77DB)

矢(U+2F6E) 矢(U+77E2)
石(U+2F6F) 石(U+77F3)
示(U+2F70) 示(U+793A)
肉(U+2F71) 肉(U+79B8)
禾(U+2F72) 禾(U+79BE)
穴(U+2F73) 穴(U+7A74)
立(U+2F74) 立(U+7ACB)
竹(U+2F75) 竹(U+7AF9)
米(U+2F76) 米(U+7C73)
糸(U+2F77) 糸(U+7CF8)
缶(U+2F78) 缶(U+7F36)
网(U+2F79) 网(U+7F51)
羊(U+2F7A) 羊(U+7F8A)
羽(U+2F7B) 羽(U+7FBD)
老(U+2F7C) 老(U+8001)
而(U+2F7D) 而(U+800C)
耒(U+2F7E) 耒(U+8012)
耳(U+2F7F) 耳(U+8033)
聿(U+2F80) 聿(U+807F)
肉(U+2F81) 肉(U+8089)
臣(U+2F82) 臣(U+81E3)
自(U+2F83) 自(U+81EA)
至(U+2F84) 至(U+81F3)
臼(U+2F85) 臼(U+81FC)
舌(U+2F86) 舌(U+820C)
舛(U+2F87) 舛(U+821B)
舟(U+2F88) 舟(U+821F)
艮(U+2F89) 艮(U+826E)
色(U+2F8A) 色(U+8272)
艸(U+2F8B) 艸(U+8278)
虍(U+2F8C) 虍(U+864D)
虫(U+2F8D) 虫(U+866B)
血(U+2F8E) 血(U+8840)
行(U+2F8F) 行(U+884C)
衣(U+2F90) 衣(U+8863)
酉(U+2F91) 酉(U+897E)
見(U+2F92) 見(U+898B)

角(U+2F93) 角(U+89D2)
言(U+2F94) 言(U+8A00)
谷(U+2F95) 谷(U+8C37)
豆(U+2F96) 豆(U+8C46)
豕(U+2F97) 豕(U+8C55)
豸(U+2F98) 豸(U+8C78)
貝(U+2F99) 貝(U+8C9D)
赤(U+2F9A) 赤(U+8D64)
走(U+2F9B) 走(U+8D70)
足(U+2F9C) 足(U+8DB3)
身(U+2F9D) 身(U+8EAB)
車(U+2F9E) 車(U+8ECA)
辛(U+2F9F) 辛(U+8F9B)
辰(U+2FA0) 辰(U+8FB0)
辵(U+2FA1) 辵(U+8FB5)
邑(U+2FA2) 邑(U+9091)
酉(U+2FA3) 酉(U+9149)
采(U+2FA4) 采(U+91C6)
里(U+2FA5) 里(U+91CC)
金(U+2FA6) 金(U+91D1)
長(U+2FA7) 長(U+9577)
門(U+2FA8) 門(U+9580)
阜(U+2FA9) 阜(U+961C)
隶(U+2FAA) 隶(U+96B6)
隹(U+2FAB) 隹(U+96B9)
雨(U+2FAC) 雨(U+96E8)
青(U+2FAD) 青(U+9751)
非(U+2FAE) 非(U+975E)
面(U+2FAF) 面(U+9762)
革(U+2FB0) 革(U+9769)
韋(U+2FB1) 韋(U+97CB)
韭(U+2FB2) 韭(U+97ED)
音(U+2FB3) 音(U+97F3)
頁(U+2FB4) 頁(U+9801)
風(U+2FB5) 風(U+98A8)
飛(U+2FB6) 飛(U+98DB)
食(U+2FB7) 食(U+98DF)

首(U+2FB8) 首(U+9996)
香(U+2FB9) 香(U+9999)
馬(U+2FBA) 馬(U+99AC)
骨(U+2FBB) 骨(U+9AA8)
高(U+2FBC) 高(U+9AD8)
髟(U+2FBD) 髟(U+9ADF)
鬥(U+2FBE) 鬥(U+9B25)
鬯(U+2FBF) 鬯(U+9B2F)
鬲(U+2FC0) 鬲(U+9B32)
鬼(U+2FC1) 鬼(U+9B3C)
魚(U+2FC2) 魚(U+9B5A)
鳥(U+2FC3) 鳥(U+9CE5)
鹵(U+2FC4) 鹵(U+9E75)
鹿(U+2FC5) 鹿(U+9E7F)
麥(U+2FC6) 麥(U+9EA5)
麻(U+2FC7) 麻(U+9EBB)
黃(U+2FC8) 黃(U+9EC3)
黍(U+2FC9) 黍(U+9ECD)
黑(U+2FCA) 黑(U+9ED1)
黻(U+2FCB) 黻(U+9EF9)
黽(U+2FCC) 黽(U+9EFD)
鼎(U+2FCD) 鼎(U+9F0E)
鼓(U+2FCE) 鼓(U+9F13)
鼠(U+2FCF) 鼠(U+9F20)
鼻(U+2FD0) 鼻(U+9F3B)
齊(U+2FD1) 齊(U+9F4A)
齒(U+2FD2) 齒(U+9F52)
龍(U+2FD3) 龍(U+9F8D)
龜(U+2FD4) 龜(U+9F9C)
龠(U+2FD5) 龠(U+9FA0)

3.2 Among CJK Radical Supplement(U+2E80 - U+2EFF), the following characters should not be used. Use the alternative characters instead.

乚(U+2E83) 乚(U+4E5A)
亅(U+2E85) 亅(U+4EBB)
冫(U+2E89) 冫(U+5202)
匚(U+2E8B) 匚(U+353E)

允(U+2E8F) 允(U+5C23)
尢(U+2E90) 尢(U+5C22)
巳(U+2E92) 巳(U+5DF3)
幺(U+2E93) 幺(U+5E7A)
互(U+2E94) 互(U+5F51)
冫(U+2E95) 冫(U+5F50)
亅(U+2E96) 亅(U+5FC4)
扌(U+2E98) 扌(U+624C)
攴(U+2E99) 攴(U+6535)
无(U+2E9B) 无(U+65E1)
𠂇(U+2E9C) 日(U+65E5)
夂(U+2E9E) 夂(U+6B7A)
母(U+2E9F) 母(U+6BCD)
民(U+2EA0) 民(U+6C11)
彡(U+2EA1) 彡(U+6C35)
氷(U+2EA2) 氷(U+6C3A)
灬(U+2EA3) 灬(U+706C)
冫(U+2EA5) 冫(U+722B)
斗(U+2EA6) 斗(U+4E2C)
𠂇(U+2EA7) 生(U+20092)
彡(U+2EA8) 彡(U+72AD)
王(U+2EA9) 王(U+738B)
𠂇(U+2EAB) 𠂇(U+7F52)
示(U+2EAC) 示(U+793A)
示(U+2EAD) 示(U+793B)
𠂇(U+2EAE) 竹(U+7AF9)
糸(U+2EAF) 糸(U+7CF9)
纟(U+2EB0) 纟(U+7E9F)
𠂇(U+2EB1) 𠂇(U+7F53)
𠂇(U+2EB2) 𠂇(U+7F52)
𠂇(U+2EB5) 𠂇(U+2626B)
𠂇(U+2EB7) 𠂇(U+2634C)
𠂇(U+2EB9) 𠂇(U+8002)
𠂇(U+2EBA) 𠂇(U+8080)
𠂇(U+2EBD) 𠂇(U+26951)
++(U+2EBE) ++(U+8279)
++(U+2EBF) ++(U+8279)

𠂇(U+2EC0) 𠂇(U+8279)
虎(U+2EC1) 虎(U+864E)
𠂈(U+2EC2) 𠂈(U+8864)
𠂉(U+2EC3) 𠂉(U+8980)
西(U+2EC4) 西(U+897F)
见(U+2EC5) 见(U+89C1)
角(U+2EC6) 角(U+89D2)
肉(U+2EC7) 𠂊(U+278B2)
𠂋(U+2EC8) 𠂋(U+8BA0)
贝(U+2EC9) 贝(U+8D1D)
车(U+2ECB) 车(U+8F66)
𠂌(U+2ECC) 𠂌(U+8FB6)
𠂍(U+2ECD) 𠂍(U+8FB6)
𠂎(U+2ECE) 𠂎(U+8FB6)
𠂏(U+2ECF) 𠂏(U+961D)
𠂐(U+2ED0) 𠂐(U+9485)
長(U+2ED1) 長(U+9577)
长(U+2ED3) 长(U+957F)
𠂑(U+2ED7) 𠂑(U+96E8)
青(U+2ED8) 青(U+9752)
韦(U+2ED9) 韦(U+97E6)
页(U+2EDA) 页(U+9875)
风(U+2EDB) 风(U+98CE)
飞(U+2EDC) 飞(U+98DE)
食(U+2EDD) 食(U+98DF)
𠂒(U+2EDE) 𠂒(U+2967F)
食(U+2EDF) 食(U+98E0)
𠂓(U+2EE0) 𠂓(U+9963)
𠂔(U+2EE1) 𠂔(U+29810)
马(U+2EE2) 马(U+9A6C)
骨(U+2EE3) 骨(U+9AA8)
鬼(U+2EE4) 鬼(U+9B3C)
鱼(U+2EE5) 鱼(U+9C7C)
鸟(U+2EE6) 鸟(U+9E1F)
麦(U+2EE8) 麦(U+9EA6)
黄(U+2EE9) 黄(U+9EC4)
𠂕(U+2EEA) 𠂕(U+9EFE)

齊(U+2EEB) 齊(U+6589)
齊(U+2EEC) 齊(U+9F50)
齒(U+2EED) 齒(U+6B6F)
齒(U+2EEE) 齒(U+9F7F)
竜(U+2EEF) 竜(U+7ADC)
龍(U+2EF0) 龍(U+9F99)
龜(U+2EF1) 龜(U+9F9C)
龜(U+2EF2) 龜(U+4E80)
龟(U+2EF3) 龟(U+2EF3)

3.3 Compatibility Characters (U+F900 to U+FAFF) should not be used. If your IDS contains compatibility characters, replace them with the corresponding CJK Ideographs.

4. Sample file.

In this archive, the sample IDS data which covers most of BMP and SIP characters are attached. They might be useful on constituting the IDS of any character, as the most efficient way to constitute the IDS is to refer to the IDS of the similar character and copy (the part of) them.

If you can't find the appropriate DC of the target character, think of any other character which shares the common DC part, then search and see how that character is constituted in the sample file.

5. Special Characters

Some ideograph components are not appeared in the UCS. Therefore, Mr. Taichi has used several specially designed ideograph components in PUA. The font called eudc.ttf is included in this archive which assigns some Ideograph glyphs in PUA.

Attached file `CDP-list.pdf(IRG N1108) describes some of these characters.