

Universal Character Set

UCS

ISO/IEC JTC1/SC2/WG2 IRG N2225Draft for review

Date: 2017-07-16

Source:	Henry Chan and Qin Lu (IRG Rapporteur)
Title:	CJK Supplementary Components for IDS Use
Distribution:	IRG Members and Ideographic Experts
Medium:	Electronic
Pages:	8

1. Introduction

Based on the IRG Recommendations IRG M48.10, a list of supplementary CJK ideograph components are listed here for use in IRG WS 2017 submissions using IDS with these supplements to increase quality of machine checking of IDS. If this is successful, IRG will consider adding this as a new IWDS series.

The components listed here are mostly extracted from analysis of CJK Ideographs in the URO. The authors consider them to be in common use, relatively speaking. To make identification easy, the components are organized according the first stroke (FS) listed below:

FS	FS Code	Name of Stroke	Rep Glyph	Character Examples
1	H	Horizontal stroke (“橫”, <i>Héng</i>)	一	一 (one), 三 (three)
2	S	Vertical stroke (“豎”, <i>Shù</i>)	丨	卜 (divination), 巾 (thirty)
3	P	Left slant stroke (“撇”, <i>Piě</i>)	丿	川 (river), 月 (moon)
4	D	Dot stroke (“點”, <i>Diǎn</i>)	丶	流 (current), 宝 (precious)
5	Z	Turned stroke (“折”, <i>Zhé</i>) ¹ .	乚 乚	𠂇, 予, 細

Suggestion: Using traditional Chinese characters for “橫”, “豎” and “點”, as in Annex K of IRG PnP.

Each component is assigned an IDS Code. The IDS code of the components is composed of the FS Code followed by the number of strokes of the component, followed by a separator “—”, followed by a serial number in two decimal digits.

¹ There are many different forms of the turned strokes and their starting stroke can be any of the above single strokes. No distinction is made in this document.

𠂔

For example, the component 𠂔 will be given the IDS code H7-01. Thus, the IDS of 曹 is defined as 曹 H7-01 𠂔.

Each component may have multiple variant glyphs. A head glyph is selected and the other variant glyphs are placed underneath it in each row. The number of strokes is based on the head glyph for the component.

2. List of Components

Below is the list of the supplementary components organized according to the first stroke.

2.1. FS = 1 (FS Code = H)

IDS code	Glyph	Description
H3-01	𠂔	CJK RADICAL PAW FOUR ² Example: bottom of 虐 / 疔 = CDP-8DC1
H3-02	𠂔	CJK COMPONENT C-SIMPLIFIED TRIPLE EARTH Example: 尧 = CDP-89D9
H3-03	𠂔	CJK COMPONENT CAI Example: outside of 在 / 存 = CDP-88F1
H3-04	𠂔	CJK COMPONENT RAISE ONE Example: bottom of 具 / 巽 / 共
H5-01	𠂔	CJK COMPONENT SACK DRAWSTRING WITH STOPPER Example: top of 囊 and 囊 = CDP-8DDD
H6-01	𠂔	CJK COMPONENT BEAST BODY TWO Example: bottom of 麤焉烏 = CDP-8DBF
H6-02	𠂔	CJK COMPONENT LIGATURE DOUBLE ARRIVE Example: top of 晉 = CDP-8DAB
H6-03	𠂔	CJK COMPONENT PART HORSE Example: outside of 馬 = CDP-896A

² The name of the components used in this document is only meant as a reference. It has no significance in the IDS sequence. So, its appropriateness will not be discussed for WS 2017. However, if the naming of these components is changed as a result of IRG/WG discussion, update to this document will be done accordingly.

H7-01	𠂇	CJK COMPONENT LIGATURE DOUBLE EAST Example: top of 曹 / 曹 = CDP-8D61
H8-01	頁	CJK COMPONENT LEAF Example: top of 𦵏 / 𦵏 = CDP-8CD4
H9-01	𠂇	CJK COMPONENT SIMPLIFIED OVERSIGHT Example: top of 監 / 覽 / 𦵏 = CDP-8CC9
H14-01	𠂇	CJK COMPONENT OVERSIGHT Example: top of 鑒 / 覽 = CDP-8D50

2.2. FS = 2 (FS Code = S)







IDS Code	Glyph	Description
S2-01	リ	CJK COMPONENT C-SIMPLIFIED MOUND Example: left of 帥 / 归 = CDP-8968
S4-01	収	CJK COMPONENT C-SIMPLIFIED LEAD Example: top of 堅 / 豎 / 紧 / 腎 / 賢 = CDPO-81EF / UTC-02964
S4-02	冂	CJK COMPONENT BOTTOMLESS HOLLOW Example: bottom of 雋 = CDP-8DDF
S5-01	𠂇	CJK COMPONENT C-SIMPLIFIED OVERSIGHT Example: top of 監 / 覽 = CDPO-81F0 / UTC-02965
S5-02	电	CJK COMPONENT REPTILE BODY Example: body of 竜, 龟 = CDP-8BE8
S7-01	𦵏	CJK COMPONENT COCOON Example: outer enclosure of 繭 / 繭
S7-02	𦵏	CJK COMPONENT VASE BOTTOM Example: bottom of 壺 / 壺 / 歡
S8-01	𦵏	CJK COMPONENT DRAGON BODY Example: right of 龍 = CDP-89B0
S9-01	𦵏	CJK RADICAL J-SIMPLIFIED FROG Example: right of 繩 / 蠅

S11-01	𡈼 𡈼 𡈼	CJK COMPONENT DREAMING PERSON Example: top of 夢 / 薨 / 薈 / 薹 / 薺 / 薻 / 薼 = CDP-8D60
--------	-------------	--






2.3. FS = 3 (FS Code = P)

IDS Code	Glyph	Description
P3-01	ㇰ	CJK RADICAL PAW THREE Example: left of 印 / 褻 / 褻 / 褻 / 褻 / 處
P3-02	ㇱ	CJK COMPONENT LEFT ENCLOSURE Example: left of 兜
P4-01	ㇲ	CJK RADICAL V-SIMPLIFIED WIND Example: right of 飢
P4-02	ㇳ	CJK RADICAL MEAT TWO Example: top of 炙 / 𪚩 / 然 / 將
P4-03	ㇴ	CJK RADICAL MOON TWO Example: top right of 望 = CDP-8B5E
P4-04	ㇵ	CJK COMPONENT LIGATURE DOUBLE HUMAN Example: bottom right of 旅 (J-source) / right of 𠂇
P4-05	ㇶ	CJK COMPONENT SIMPLIFIED PART BIRD Example: outside of 鸟 / 岛
P5-01	ㇷ	CJK COMPONENT LIU Example: top of 留 / top left of 劉 = CDP-8C69
P5-02	ㇸ	CJK COMPONENT LOOSE ROCK ONE Example: left of 段 = CDP-8CE4
P6-01	ㇹ	CJK COMPONENT BEAST BODY ONE Example: bottom of 𪚩 = CDP-8C40
P6-02	ㇺ	CJK COMPONENT BEAST HEAD ONE Example: top of 象 / 𪚩
P6-03	ㇻ 𪚩 𪚩	CJK COMPONENT COMPLEX PERSON CAVE Example: top of 𪚩 / 𪚩 = CDP-8CAF

Should the number of strokes be 2 instead of 3?

P7-01		CJK COMPONENT FOOD VESSEL TWO Example: left of 即 / 既	The left part of the example “既” is different from the glyph in question. Shall we change it to “既” (U+65E3)?
P7-02		CJK COMPONENT BEAST BODY THREE Example: bottom of 𪛗 / 𪛘 / 𪛙 / 𪛚 / 𪛛 / 𪛜	
P7-03		CJK COMPONENT PART BIRD Example: outside of 島 / 鳥 = CDP-8CBB	
P8-01		CJK COMPONENT DANCE Example: top of 無 / 舞 = CDP-8BBE	
P10-01		CJK COMPONENT RAISE TWO Example: enclosure of 與 / 與 / 興 / 興 = CDP-8BA8	
P15-01		CJK COMPONENT COMPLEX POURING VESSEL Example: top of 𪛗 / 𪛘	

2.4. FS = 4 (FS Code = D)

IDS Code	Glyph	Description	
D4-01		CJK COMPONENT OMISSION Example: bottom of 𪛗 / bottom right of 𪛘 = CDP-89AE	
D5-02		CJK COMPONENT C-SIMPLIFIED LEARN Example: top of 学 / 𪛗 / 𪛘 / 𪛙 / 𪛚 / 𪛛 / 𪛜 etc. = CDPO-81DD / UTC-02966	
D6-01		CJK COMPONENT C-SIMPLIFIED CHAOS Example: top of 恋 / 𪛗 / 𪛘. Note: use 4EA6 亦 instead where appropriate, such as 5955 𪛗 = UTC-02967	
D6-02		CJK COMPONENT VIRTUOUS Example: left of 𪛗 / 朗, left hand side form of 良 = CDP-8C42	
D6-03		CJK COMPONENT BEAST HEAD TWO Example: outside of 鹿 / 𪛗 = CDP-8D56	Should the number of strokes be 7 instead of 6?

D7-01	宀	CJK COMPONENT ARCHITECTURAL STRUCTURE Example: top of 亮 / 亭 = CDP-8C4D
D7-02	宀	CJK COMPONENT GUEST Example: top of 賓 = CDP-8D74
D10-01	廴	CJK COMPONENT EVEN Example: top of 壘 / 齊 = CDP-8DEB

2.5. FS = 5 (FS Code = Z)

IDS Code	Glyph	Description
Z2-01	勹	CJK COMPONENT RIGHT ENCLOSURE Example: right of 兜 / top right of 段
Z2-02	勹	CJK COMPONENT LEFT HAND ONE Example: left of 収
Z2-03	乚	CJK COMPONENT LEFT HAND TWO Example: right of 𠂔 = CDP-88B5
Z2-04	𠂔	CJK COMPONENT SIMPLIFIED PART HORSE Example: outside of 马
Z2-05	𠂔	CJK COMPONENT RIGHT HAND TWO Example: right top of 祭 / 登 = CDP-8C4E
Z3-01	𠂔	CJK COMPONENT RIGHT HAND ONE Example: top of 尋 / top right of 侵 / bottom of 𢇛 = CDP-8BAB
Z3-02	𠂔	CJK COMPONENT HOLD ONE Example: right top of 恐 / 鞏 etc
Z3-03 ³	丸	CJK COMPONENT HOLD TWO (variant of 𠂔) Example: right of 執 / 執 / 狢 / 秣 Note: use 4E38 丸 (wan2) where appropriate, e.g. 7D08 紈.
Z4-01	𠂔	CJK COMPONENT TILTED HEAD Example: bottom of 吳 = CDP-8CD5
Z4-02	丹	CJK COMPONENT RAN Example: left of 那 = CDP-89FA

³ Refer to Appendix A: Rationale for Inclusion of Z3-03.

Z5-01	艮	CJK COMPONENT FOOD VESSEL ONE Example: left of 即 / 既 = CDP-8B7C
Z5-02	𠂔	CJK COMPONENT LOOSE ROCK TWO Example: left of 段 = CDP-8C7A
Z7-01	𠂔	CJK COMPONENT BEAST BODY THREE Example: bottom of 鼠 / 鼯 = CDP-8D46

3. Additional Components for Reference

The table below also supply some coded ideographs which can serve as common component variants. For easy IDS checking, the base character as a component should be used in the IDS sequence to represent the variant glyph directly.

Base Character	Variant Glyph	Descriptions
U+2B94E 𠂔	𠂔	Example: right of 𠂔 / 𠂔
U+53C0 𠂔	𠂔	Example: top of 𠂔 / 𠂔 = CDP-8BD0
U+5C6F 屯	𠂔	Example: right of 𠂔 / 𠂔 / 𠂔 = CDP-8AF0
U+6B62 止	𠂔	Example: bottom of 𠂔 / 𠂔 = CDP-8962
U+9FB0 𠂔	𠂔	Example: bottom of 𠂔 / 𠂔 / 𠂔 = CDP-8961
U+3B30 𠂔	𠂔	Example: right of 𠂔 / 𠂔 / 𠂔 = CDP-89E7
U+4E23 𠂔	𠂔	Example: top right of 𠂔, right of 𠂔 / 𠂔 / 𠂔 = CDPO-85FD

Appendix A: Rationale for Inclusion of Z3-03

Z3-03 丸 is included as a separate component from U+4E38 丸. First, they are considered different components in the Taiwan and Hong Kong locale. For characters whose etymology is U+4E2E 𢇛 (to hold), the glyph shape Z3-03 with the dot protruding is used. For characters whose etymology is U+4E38 丸 (small round object), the T/H glyph shape of U+4E38 with the dot not protruding is used.

This is a similar case to the existing dis-unified radicals 2EC0 卄 CJK RADICAL GRASS THREE from 2EBE 卅 CJK RADICAL GRASS ONE, and 2EBC 月 CJK RADICAL MEAT from 2F49 月 KANGXI RADICAL MOON. The difference in the shape is reflective of a different etymology and hence “non-cognateness” of characters in the Taiwan and the Hong Kong locales.

Second, these normally unifiable differences in components could also reflect a difference in meaning in certain context for other locales, and not just as a simple typographical variation. For example, PRC selectively applies such typographical variations as a standardized difference for a range of non-cognate characters. The difference between U+2EC0(卄) and U+2EBE(卅) is used to distinguish between the two characters U+830D 苟 and U+82DF 苟 in the PRC locale.

Thus, the inclusion of Z3-03 component is also meaningful and useful for IDS composition.