

Source:	Yi BAI
Title:	Feedback on HKSCS IDS sequence
Status:	Individual Contribution on IRG #57
Action:	To be considered by IRG

1. New features of HKSCS IDS sequences

1.1 Alternative sequences

When a character can be described by more than one IDS sequence, it's obvious that a shorter IDS, or an IDS with more common characters, is preferred. However there is a trade off between a shorter IDS with uncommon characters and a longer IDS with common characters. HKSCS IDS gives both IDSeS, with the former one as primary IDS, and latter one as alternative IDS listed in "Remarks" column.

1.2 Components not to be further decomposed

Another common problem about IDS sequences is "when to stop". Lacking a clear rule will result in decomposition of very common radicals into too detailed strokes. Defining a small set of atomic radicals and characters is helpful. HKSCS IDS defines 347 radicals and characters as atomic, even some of them like 自 can be further broke into 𠄎 目, because they are atomic for common ideographs users.

1.3 Merge and separation of similar components

HKSCS IDS uses non-cognate rules to decide whether to merge or separate similar components. For example, 月 moon/月 meat pair and 月 meat/冂 hat pair are used independently, while some radical variants like 羊,牛,子(with etymological relation with 子) are merged into 羊,牛,子 respectively, to give the sequences more "semantic" meaning.

1.4 Standardized overlapping structure

Overlaid IDC 𠄎 are often abused to represent some uncommon structure, and choice of component A and B used in 𠄎AB are mostly up to author's preferences. HKSCS IDS standardized overlapping structure by defining A and B must be listed in stroke order. This partially solves the chaos in using overlapping structure.

2. Remarks about HKCSC IDS data

(a) Possible better IDS

	Current IDS	Proposed IDS	Note
辭	𠄎 𠄎 𠄎 𠄎 自 辛	𠄎 𠄎 辛	
僵	𠄎 𠄎 𠄎 𠄎 𠄎 𠄎	𠄎 𠄎 𠄎 𠄎 𠄎	more accurate
壻	𠄎 土 胥	𠄎 土 胥	semantic part is "土"

携	扌 山隹	扌 隹	
撮	扌 旦耳彡	扌 月一耳彡	right-upper part is not 旦
滔	氵 𠂔缶	氵 缶	G-glyph of 缶 exact matches right part
瑶	王 𠂔缶	王 缶	G-glyph of 缶 exact matches right part
曦	日 羊 秀 戈	日 羲	T-glyph of 羲 exact matches right part
榷	木 𠂔 一 佳	木 榷	except dot replaced with left-falling stroke, 榷 matches right part exactly
繫	纟 业 白 羊 彡 米	纟 穀 米	
絲	纟 幺 幺 八 八	纟 幺 八 八	
蠹	土 𠂔 石 虫	土 虫	
惠	一 由 心	由 心	
蒟	甘 巾 入 入	甘 巾 从	
角	夕 冂 欠	夕 冂	
灑	氵 勑 鳥	氵 灑	
汞	丩 水	汞	
廉	宀 冫 厂 邑 卅 八 丨 丨	产 邑 卅 八 丨 丨	
稟	宀 回 示	稟 示	
藁	土 𠂔 禾	土 禾	
壘	方 方 方 土	方 壘	
漫	艹 氵 帚 又	艹 漫	except right-upper part, G-glyph of 漫 exactly match lower part

(b) Possible additional alternative IDses in "Remarks" column

	Current IDS	Proposed alternative IDS	Note
敷	虎 魚 女	廌 女	
懷	忄 𠂔 衣	忄 𠂔 冫	
炼	火 东 一	火 东	

爪	ㇰ一ㇰ儿	ㇰ一爪	
衛	ㇰ彳 ㇰZH3口巾子	ㇰ彳 衛子	
奥	ㇰ丿 ㇰ冂米大	ㇰ丿 奥大	
荦	ㇰ艹ㇰ牛	ㇰ荦牛	
莹	ㇰ艹ㇰ玉	ㇰ荦玉	
萦	ㇰ艹ㇰ糸	ㇰ荦糸	
萤	ㇰ艹ㇰ虫	ㇰ荦虫	
莖	ㇰ艹ㇰ金	ㇰ荦金	

(c) Possible additional alternative IDSes in "Remarks" column, if Extension G characters can be used

	Current IDS	Proposed alternative IDS	Note
親	ㇰㇰ土示見	ㇰ亲見	
擧	ㇰㇰ臣ㇰㇰㇰ一ㇰ手	ㇰ臨手	
画	ㇰ一ㇰ口田	ㇰ一画	
皋	ㇰ臭ㇰ二	ㇰ臭丰	
癩	ㇰ疒ㇰ乡佳	ㇰ疒維	
奥	ㇰ丿 ㇰ冂米大	ㇰ丿 奥大	
变	ㇰ亦又	ㇰ亦又	亦 is non-cognate with 亦
栾	ㇰ亦木	ㇰ亦木	
脍	ㇰ亦肉	ㇰ亦肉	
銮	ㇰ亦金	ㇰ亦金	
鳶	ㇰ亦鳥	ㇰ亦鳥	
学	ㇰㇰㇰ子	ㇰ学子	
荣	ㇰㇰㇰ木	ㇰ荣木	

(d) other issues

Components in overlapping structures didn't follow the "stroke order" rule. However, there seems no better way to break down them into 2 components.

Character	IDS
斲	ㇰㇰㇰㇰ王口斤
嫁	ㇰ女ㇰ象、

(End of document).