中文界面諮詢委員會

中文資訊科技工作小組

進展報告 1

1999.8.11

1. 工作小組工作範圍（中英文版）

中文版建議爲
- 就採用 ISO 10646 國際編碼標準的技術問題提供意見；
- 促使本地軟硬件商開發採用 ISO 10646 國際編碼標準的中文應用系統；
- 推廣採用 ISO 10646 國際編碼標準的中文應用系統。

英文版建議爲
- to advise on the technical issues arising from the adoption of ISO 10646;
- to facilitate local developers in the development of ISO 10646 enabled Chinese applications;
- to promote the use of ISO 10646 enabled Chinese applications.

2. 「香港增補字符集」Big5 編碼
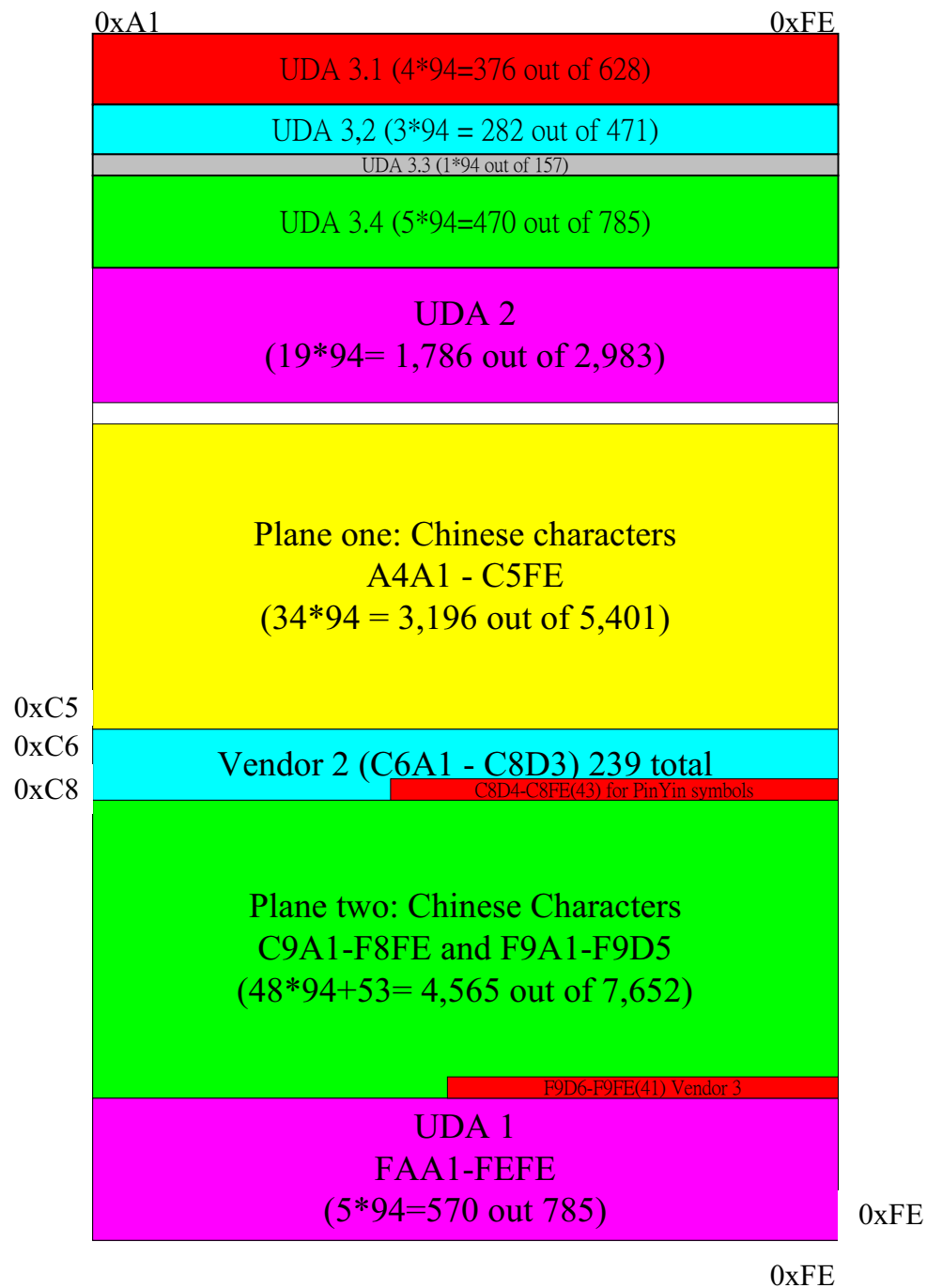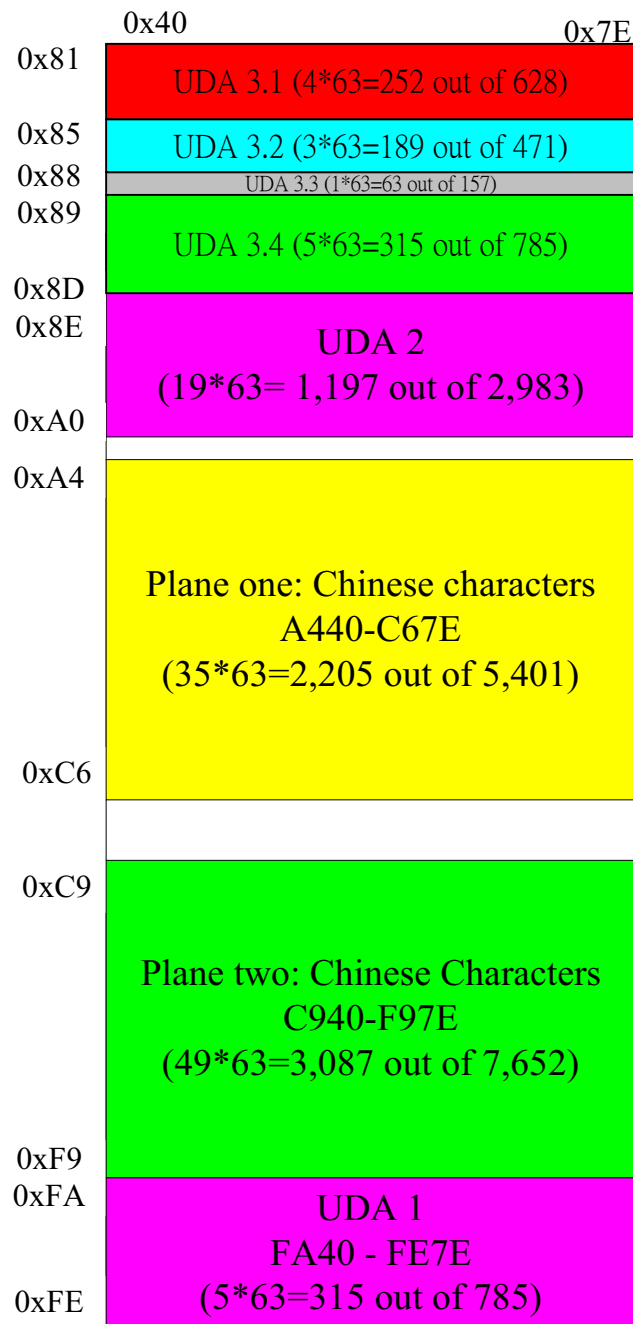
討論結果如附錄 1 所載。

3. 「香港增補字符集」Unicode 編碼

討論結果如附錄 2 所載。

Big5 Code Ranges & Suggested Allocation of Hong Kong Supplementary Character Set (HKSCS)

1. Big5 is a fixed 2-octet scheme.   The first octet ranges from 0x81 to 0xFE and the second octet with two separate code ranges are valued   from 0x40 to 0x7E and 0xA1 to 0xFE, respectively.   Thus the diagram shows the second octet with two disjoint ranges.

Range (hex)	Total No. of code points	Name of area	Secondary range (total code points)	Currently used for	Suggested use	Remarks
8140 - 8D7E 81A1 - 8DFE	2,041	UDA 3	UDA3.1: 8140 - 847E(252) 81A1 - 84FE (376) Total: 628	No	UDA3.1: HKSCS future extension will not be using this area. It is suggested for private use UDA3.2: HKSCS future extension on Chinese characters. It is intended to extend downwards(start from the highest row towards lower numbered rows). UDA3.3: HKSCS symbol zone and HKSCS future extension	
			UDA3.2: 8540 - 877E(189) 85A1 -87FE (282) Total: 471			
			UDA3.3 8840 – 887E(63) 88A1 – 88FE (94) Total: 157			
			UDA3.4: 8940 - 8D7E(315) 89A1-8DFE(470) Total: 785	Mainly used by Judiciary systems in HKSAR	Reserve for HKSCS and future expansion	Currently has 166 code points free for HKSCS extension

8E40 - A07E 8EA1- A0FE	2,983	UDA 2		Occupied by HKSCS already	No	
A140 - A37E A1A1 - A3FE	471			Big5 symbols	Not suggested for use by HKSCS	
A440 - C67E A4A1 - C5FE	5,401	Big5 primary set(Plane 1)		Big5 primary set	No	
C740 - C87E	126			Some vendor use	Not suggested for use by HKSCS	
C6A1 - C8FE	282	Vendor 2	C6A1-C8D3(239)	Vendor symbols	Suggest to keep all vendor symbols	See Eten symbols
			C8D4 - C8FE(43)	Not being used	Pin Yin and/or other symbols in HKSCS	
C940 - F97E C9A1 - F9D5	7,652	Big5 secondary set(Plane 2)		Big5 secondary set	No	
F9D6 - F9FE	41	Vendor 3 Vendor 3	F9D6 - F9DC(7)	Chinese characters missed by Big5	No	
			F9DD - F9FE(34)	Border symbols	HKSCS future symbols if needed	See Eten table
FA40 - FE7E FAA1 - FEFE	785	UDA 1		HKSCS	No	

Left diagram (0x40 – 0x7E):

0x40 ... 0x7E

0x81 — UDA 3.1 (4*63=252 out of 628)

0x85 — UDA 3.2 (3*63=189 out of 471)

0x88 — UDA 3.3 (1*63=63 out of 157)

0x89 — UDA 3.4 (5*63=315 out of 785)

0x8D

0x8E — UDA 2 (19*63= 1,197 out of 2,983)

0xA0

0xA4 — Plane one: Chinese characters A440-C67E (35*63=2,205 out of 5,401)

0xC6

0xC9 — Plane two: Chinese Characters C940-F97E (49*63=3,087 out of 7,652)

0xF9

0xFA — UDA 1 FA40 - FE7E (5*63=315 out of 785)

0xFE

Right diagram (0xA1 – 0xFE):

0xA1 ... 0xFE

UDA 3.1 (4*94=376 out of 628)

UDA 3,2 (3*94 = 282 out of 471)

UDA 3.3 (1*94 out of 157)

UDA 3.4 (5*94=470 out of 785)

UDA 2 (19*94= 1,786 out of 2,983)

Plane one: Chinese characters A4A1 - C5FE (34*94 = 3,196 out of 5,401)

0xC5

0xC6 — Vendor 2 (C6A1 - C8D3) 239 total

0xC8 — C8D4-C8FE(43) for PinYin symbols

Plane two: Chinese Characters C9A1-F8FE and F9A1-F9D5 (48*94+53= 4,565 out of 7,652)

F9D6-F9FE(41) Vendor 3

UDA 1 FAA1-FEFE (5*94=570 out 785)

0xFE

0xFE

**Recommended HKSCS (Hong Kong Supplementary Character Set) Coding Scheme
under ISO 10646/Unicode 3.0[1]**

The recommendations can be summarized as follows:

1. HKSCS characters that can be found in Unicode 3.0 will not be allocated a code point in Unicode 3.0's Private User Area (PUA).   Mapping to Unicode 3.0 will be provided in HKSCS.

2. For those characters in HKSCS which do not have matching characters in Unicode 3.0, code points will be assigned in the PUA of Unicode 3.0 and the assignment will follow the sequence of HKSCS code point allocation under BIG5 Scheme.

3. For areas reserved for HKSCS future extension under the BIG5 scheme, the same number of free code points will be reserved in Unicode 3.0's PUA.

The following is a block diagram of Unicode 3.0's PUA under the suggested code allocation scheme :

HKSCS characters and free space in UDA 3.2*   [As at August 1999, UDA 3.2 is an empty zone reserved for future expansion of HKSCS. Therefore, there will not be any character in it at the moment.]
HKSCS symbols and free space in UDA 3.3* less those that already exist in Unicode 3.0
HKSCS characters and free space in UDA 3.4* less those that already exist in Unicode 3.0
HKSCS characters and free space in UDA 2 less those that already exist in Unicode 3.0
HKSCS characters and free space in Vendor 2 less those that already exist in Unicode 3.0
HKSCS characters and free space in Vendor 3 less those that already exist in Unicode 3.0
HKSCS characters and free space in UDA 1 less those that already exist in Unicode 3.0
Genuine Unicode 3.0 PUA; HKSCS future extension (under Big5 Scheme) will not be using this area under Unicode 3.0

Note : * Please refer to the document " Big5 Code Ranges & Suggested Allocation of Hong Kong Supplementary Character Set (HKSCS)" for the definition of UDA 3.2, 3.3, 3.4.

---

[1] Unicode 3.0 is an implementation of ISO 10646 which will support all characters and code points in ISO 10646v1 and other characters approved by ISO in the next version of ISO 10646 including extension A as well as all characters found in Unicode 2.0.   Unicode 3.0 will be launched in September 1999.

Row No.:

0x00

0x00	Alphabet sets (HKSCS characters: 124 characters)
0x20	Symbols (HKSCS characters: 101)
0x30	CJK Symbols, etc (HKSCS characters: 177)
0x34	Extension A (HKSCS characters: 497)
0x4E	CJK (HKSCS characters: 2,147)
0xA0	

0xE0	Private Use Area (HKSCS characters: 1,655) (HKSCS used area: 0xE000 to EBD5)
0xEB	
0xEC	Private Use Area
0xF9	

0xFA	Compatibility Zone (HKSCS characters: 5 characters)
0xFF	