

討論新字符集字符增收程序和原則

背景

《香港增補字符集》在 1999 年 9 月出版後，市民及政府部門仍不時需要把新字符納入字符集內。為此，政府與中文界面諮詢委員會（中諮會）於 2000 年 4 月公布《香港增補字符集》字符增收程序和原則，並於 2009 年 9 月加以修訂，全面採用 ISO/IEC 10646 國際編碼標準編配的碼位。為配合新字符集（暫名為《香港字符集》）的草擬工作，有需要一併修訂字符增收程序和原則。

建議

2. 《香港增補字符集》出版多年，一向廣為市民和業界接受；優化其字符增收程序和原則，在 ISO/IEC 10646 國際編碼標準內增補香港所需的字符和字形及相關資料，也可以配合橫向擴展的工作。因此，我們建議沿用《香港增補字符集》的品牌，令市民和業界更容易採納和認同。附件一對比優化《香港增補字符集》的方案（簡稱《香港增補字符集》新版本）和《香港字符集》的異同，並列出字符增收工作的考慮因素，以供參考。

3. 附件二和附件三分別提出《香港增補字符集》新版本／《香港字符集》字符增收程序和原則的一些要點，以供討論。附件四表列 268 個曾申請增收但不獲納入《香港增補字符集》的非大五碼或《香港增補字符集－2008》字符。

徵詢意見

4. 請工作小組成員就有關建議發表意見。秘書處整理後會提交中諮會審議。

2015 年 9 月

附件一：《香港增補字符集》新版本和《香港字符集》的異同

	《香港增補字符集》新版本	《香港字符集》	備註
功用	增補 ISO/IEC 10646 國際編碼標準，加入香港字符和字形，以及《香港增補字符集－2008》的修訂。		須修訂字符增收原則予以配合。增收字符的審核工作預計會較為複雜。
字數	5009+ 約 260 字	約 18,000 字	處理字符集相關參考資料的工作量會增加。 軟件開發商須檢視現時支援《香港增補字符集》的軟件是否需要更新才能支援《香港字符集》。
編碼技術	ISO/IEC 10646 國際編碼標準字集內碼。		現行的平台和軟件都支援 ISO/IEC 10646 國際編碼標準，不須考慮由 ISO/IEC 10646 下向兼容大五碼的技術問題。
字形規範	為 ISO/IEC 10646 國際編碼標準字集內適用於香港的漢字加入香港字形。		要確定字形是否完全符合規範，需要密集人手檢定。

	《香港增補字符集》新版本	《香港字符集》	備註
字符增收原則	可考慮把字符增收原則納入文件內。		目前審核增收字符時，已檢查ISO/IEC 10646國際編碼標準字集是否已收錄該字，但預計接受橫向擴展的申請後，審核增收字符的工作量會增加。

附件二：擬議新字符集字符增收程序

由於《香港增補字符集》新版本／《香港字符集》將會增收 ISO/IEC 10646 國際編碼標準字集的字符，並以橫向擴展方式加入香港字形，預計修訂會較《香港增補字符集—2008》為多。為了日後能夠更有效增收字符，可考慮以修訂版本的方式，把增收字符正式納入《香港增補字符集》新版本／《香港字符集》，之後再按增收字符情況，把若干修訂版本合併在該字符集內。現建議大致沿用《香港增補字符集—2008》增收程序，修訂部分以藍色顯示如下：

1. 市民及政府部門可向中文界面諮詢委員會（中諮會）秘書處提交字符增收申請。申請必須連同字符的字形、讀音及意義等資料提交，最好還包括字符的用途、字源和例句。
2. 市民提交的字符，會經中諮會秘書處轉交中文界面諮詢委員會工作小組（工作小組）審核。政府部門提交的字符，會由政府內部專責小組初步審議，然後再提交工作小組審核。
3. 工作小組會根據下文的原則審核增收字符的申請。為了避免浪費資源及加快審核字符的過程，市民及政府部門應先參考字符增收原則，以免提交不必要的申請。
4. 通過工作小組審核的申請會提交中諮會確認。如申請符合“迫切需要編碼的字符”條件，中諮會秘書處會為字符編配“來源參考”，格式為 HC-9999，供申請人參考。字符的字形及相關資料會在互聯網上公布，但必須強調，這些增收字符暫時只供參考，稍後才會正式納入《香港增補字符集》新版本／《香港字符集》（見下文第 6 段）。
5. 如提交的字符不獲接納，申請人可提供進一步資料，要求工作小組重新考慮。由於這些字符大部分已編有 ISO/IEC 10646 碼位，使用者可以在支援 ISO/IEC 10646 或統一碼 (Unicode) 的平台上處理這些字符。

《香港增補字符集》新版本／《香港字符集》的正式修訂

6. 由於軟件商開發軟件支援增收字符需時，加上字符集更新過於頻密會對使用者造成不便，甚至引起混亂，工作小組會視乎實際需要，在適當時間向中諮會建議公布《香港增補字符集》新版本／《香港字符集》修訂版本，把增收字符正式納入字符集，然後再按增收字符情況，把若干修訂版本合併在該字符集內。

7. 所有通過審核但尚未納入 ISO/IEC 10646 的增收字符，均會提交國際標準化組織轄下的表意文字小組。

~~8. 為進一步與國際標準接軌，中諮會決定由二零零九年九月起，《香港增補字符集》全面採用國際標準化組織接納字符申請時所編配的 ISO/IEC 10646 正式碼位，不再另外加配香港增補字符“標準內碼”，以免字符重複編碼。以往已經公布“標準內碼”的字符，其“標準內碼”仍然生效。~~

附件三：擬議新字符集字符增收原則

《香港增補字符集》新版本／《香港字符集》字符的增收原則，除了須維持增補 ISO/IEC 10646 國際編碼標準字集的功能外，還須兼顧在 ISO/IEC 10646 國際編碼標準字集內以橫向擴展形式加入香港字形的需要。現建議編訂《香港增補字符集》新版本／《香港字符集》字符增收原則如下：

1. 《香港增補字符集》新版本／《香港字符集》只收納市民和政府部門在進行電子通訊時有需要使用的字符，增收字符僅限下列範圍：
 - 1.1. 香港社羣用字，包括：
 - (a) 粵語字（須提供依據）；
 - (b) 能證明此前已流通的專有名詞用字（須提供讀音）；以及
 - (c) 作專有名詞用的字而該字見於大型字典（包括《康熙字典》、《漢語大字典》等）。
 - 1.2. 華人社會中非地區性的新生漢字（例如：新生科學名詞用字）。
2. 申請增收的字符須按“字符等同原則”甄別。凡在《香港增補字符集》新版本／《香港字符集》內已有可視為等同之字符者(#1)，概不增收。字符等同原則指 ISO/IEC 10646 等同原則，詳見 ISO/IEC 10646: Annex S - Procedure for the unification and arrangement of CJK Ideographs。等同字符的例子有：青—青、俞—俞、直—直、并—并、植—植、餅—餅。
3. 若申請增收的字符在《香港增補字符集》新版本／《香港字符集》並無等同字符，但在 ISO/IEC 10646 國際編碼標準字集內卻有可視為等同者，會根據下列因素考慮是否納入《香港增補字符集》新版本／《香港字符集》：
 - (a) 該字符的字形是否在香港廣為接納；
 - (b) 該字符是否容易與《香港增補字符集》新版本／《香港字符集》內的字符混淆(#2)；以及

(c) 該字符是否《香港增補字符集》新版本／《香港字符集》字符的異體字(#3)。

4. 字符納入《香港增補字符集》新版本／《香港字符集》後，將會加入 ISO/IEC 10646 國際編碼標準字集內，作為香港字形。

在上述範圍以外，其他有助於電子通訊的字符，會個別考慮。

#1 7DF4 是大五碼字符，而申請增收的字符字形與 J0-4E7D 相同，申請將不獲接納。類似的例子有 7E09 的 G1-6746 字形。

7DF4
系 120.9
練 練 練 練 練 練
G1-4137 HB1-BD6D T1-6E38 J0-4E7D K0-5623 V1-627C (曾接獲查詢的字符)

7E09
系 120.10
縉 縉 縉 縉 縉
G1-6746 HB1-BFA7 T1-716F J0-6556 K0-725B (不獲增收的字符)

#2 例如非大五碼或《香港增補字符集—2008》字符 5897 (增)，其字形與大五碼字符 589E (增) 相似。類似的例子有 7523 (產) 和 7522 (產)。

5897
± 32.11
增 增 增
GE-2625 T3-4640 J0-417D (不獲增收的字符)

589E
± 32.12
增 增 增 增 增 增
G0-5476 HB1-BC57 T1-6C41 J3-2F5D K0-7172 V1-5129 (大五碼字符)

7523
生 100.6
產 產 產 產
G1-327A T3-3745 J0-3B3A K0-5F27 V1-5F47 (不獲增收的字符)

7522
生 100.6

產 產 產
GE-3372 HB1-B2A3 T1-5C28

(大五碼字符)

#3 例如非大五碼或《香港增補字符集—2008》字符 596C (獎)。
根據網頁 <http://dict2.variants.moe.edu.tw/variants/> 資料，596C與大
五碼字符 734E (獎) 是一對異體字，其中 734E 是正字。

596C
大 37.11

獎
G1-3D31

獎
T3-4648

獎
J0-547D

獎
K0-6D61

獎
V1-5159

(不獲增收的字符)

734E
犬 94.11

獎
GE-332A

獎
HB1-BCFA

獎
T1-6D64

獎
J0-6052

獎
K2-4731

(大五碼字符)

參考資料：現行的《香港增補字符集》字符增收原則

《香港增補字符集》字符增收原則

1. 凡在ISO/IEC 10646國際編碼標準字集及《香港增補字符集》內已有可視為等同之字符者，概不增收；因此，增收字符須經“字符等同原則”甄別。根據ISO/IEC 10646等同原則(即ISO/IEC 10646 “Annex S – [Procedure for the unification and arrangement of CJK Ideographs](#)”)認為等同之字符，均視為等同(例如：青 — 青、兪 — 兪、直 — 直、并 — 并、植 — 植、餅 — 餅)。

2. 字符集只收納市民和政府部門在進行電子通訊時有需要使用的字符，增收字符僅限下列範圍：

2.1. 香港社羣用字，包括：

- (a) 粵語字 (須提供依據)；
- (b) 能證明此前已流通的專有名詞用字(須提供讀音)；以及
- (c) 作專有名詞用的字而該字見於大型字典(包括《康熙字典》、《漢語大字典》等)者。

2.2. 華人社會中非地區性的新生漢字(例如：新生科學名詞用字)。

在上述範圍以外，其他有助於電子通訊的字符，會個別考慮。

注：查核字符是否已收錄於 ISO/IEC 10646 國際編碼標準字集及《香港增補字符集》的方法見附件。

(修訂日期 10.2011)

附件四：不獲納入《香港增補字符集》的字符

由 2008 年 3 月 31 日至 2015 年 8 月 21 日共處理了 908 字符，全部無須納入《香港增補字符集》，當中 268 個並不屬大五碼或《香港增補字符集－2008》字符，現表列如下：

序號	碼位	字形 (#1)	提交字形
1	343C	併	併
2	3463	侏	侏
3	3484	儼	儼
4	3500	剽	剽
5	3566	吳	吳
6	36DC	嫫	嫫
7	3769	康	康
8	37A9	崑	崑
9	37B6	岐	岐
10	37D3	嶠	嶠

序號	碼位	字形 (#1)	提交字形
11	382F	呂	呂
12	3867	曠	曠
13	3B79	椏	椏
14	3BCB	穎	穎
15	3E0C	燿	燿
16	3EAC	玗	玗
17	3F66	罨	罨
18	4016	眅	眅
19	402A	眈	眈
20	4112	袞	袞

序號	碼位	字形 (#1)	提交字形
21	4120	禮	禮
22	4170	穗	穗
23	4184	禮	禮
24	41F8	篁	篁
25	42A5	糲	糲
26	42DA	綉	綉
27	4343	畚	畚
28	44C2	菩	菩
29	4591	彪	彪
30	46E5	諦	諦
31	4777	賀	賀
32	4802	楚	楚

序號	碼位	字形 (#1)	提交字形
33	48BF	郊	郊
34	4961	鏢	鏢
35	4969	鏢	鏢
36	4A1E	靄	靄
37	4AB0	靄	音意
38	4B9E	驕	驕
39	4BEA	顛	顛
40	4BEB	騎	騎
41	4E17	卅	卅
42	4E51	彳	彳
43	4EE2	彳	彳
44	4EF9	伴	伴

序號	碼位	字形 (#1)	提交字形
45	4F06	仸	仸
46	4F40	侶	侶
47	4F71	仝	仝
48	4F8E	侏	侏
49	4FB1	侏	侏
50	5038	侏	侏
51	5059	侏	侏
52	5086	侏	侏
53	50F7	儻	儻
54	51D6	準	準
55	51E6	廸	廸
56	51EA	風	風

序號	碼位	字形 (#1)	提交字形
57	520F	劊	劊
58	5215	劊	劊
59	52C6	劊	劊
60	52F2	勳	勳
61	534B	尗	尗
62	53A1	原	原
63	53B7	尗	尗
64	53C4	叁	叁
65	53DC	麥	麥
66	53DE	馭	馭
67	548A	味	味
68	5568	晴	晴

序號	碼位	字形 (#1)	提交字形
69	5585	啞	啞
70	561A	嘒	嘒
71	56E6	囧	囧
72	56F8	囧	囧
73	5717	圖	圖
74	57A6	垦	垦
75	57CA	峯	峯
76	5867	塏	塏
77	5897	增	增
78	58AD	塏	塏
79	58C4	塏	塏
80	58E1	齧	齧

序號	碼位	字形 (#1)	提交字形
81	5933	忝	忝
82	596C	獎	獎
83	598C	姘	姘
84	5A6F	嬖	嬖
85	5B04	嬖	嬖
86	5B5E	忝	忝
87	5B94	宝	宝
88	5BDC	寧	寧
89	5BFD	孚	孚
90	5C02	專	專
91	5C17	尗	尗
92	5C19	尗	尗

序號	碼位	字形 (#1)	提交字形
93	5C1B	尛	尛
94	5C84	岫	岫
95	5CB9	岯	岯
96	5CDC	岵	岵
97	5CF3	岷	岷
98	5D0A	岽	岽
99	5DBE	岿	岿
100	5EF0	廳	廳
101	6071	悅	悅
102	6074	怠	怠
103	60A6	悦	悦
104	61DC	憊	憊

序號	碼位	字形 (#1)	提交字形
105	62A5	报	报
106	63A6	揚	揚
107	6483	擊	擊
108	653C	攷	攷
109	654E	教	教
110	662A	昇	昇
111	6660	臧	臧
112	6675	啓	啓
113	669C	晉	晉
114	66D0	壘	壘
115	66D4	瞰	瞰
116	6724	鬲	鬲

序號	碼位	字形 (#1)	提交字形
117	674A	柎	柎
118	67BC	葉	葉
119	6808	棧	棧
120	68C1	稅	稅
121	68DB	梢	梢
122	68FF	梲	梲
123	698F	楹	楹
124	6B1B	欂	欂
125	6C44	汜	汜
126	6C56	冱	冱
127	6C63	次	次
128	6C91	汙	汙

序號	碼位	字形 (#1)	提交字形
129	6CA1	沒	沒
130	6D23	泚	泚
131	6D81	泚	泚
132	6E42	澇	澇
133	6F71	澆	澆
134	7025	澇	澇
135	702E	灑	灑
136	7036	灑	灑
137	708C	炔	炔
138	7097	炔	炔
139	70B2	焔	焔
140	7103	焔	焔

序號	碼位	字形 (#1)	提交字形
141	712D	堯	堯
142	7138	熨	熨
143	7148	颯	颯
144	717F	燻	燻
145	718B	夔	夔
146	7271	犊	犊
147	7390	玦	玦
148	743B	璿	璿
149	743E	璿	璿
150	7484	境	境
151	7520	眚	眚
152	7523	產	產

序號	碼位	字形 (#1)	提交字形
153	7698	晴	晴
154	76A2	曉	曉
155	76A9	覬	覬
156	76C1	盈	盈
157	76D8	盘	盘
158	7700	明	明
159	77AE	瞰	瞰
160	7807	砭	砭
161	780B	砭	砭
162	7872	砭	砭
163	7892	砭	砭
164	7907	礪	礪

序號	碼位	字形 (#1)	提交字形
165	79B5	颯	颯
166	7A0F	稈	稈
167	7A16	稭	稭
168	7A33	稔	稔
169	7A42	穗	穗
170	7AB9	籟	籟
171	7AE1	苜	苜
172	7AF8	競	競
173	7B57	筭	筭
174	7D2D	紂	紂
175	7DC3	縱	縱
176	7DC8	緯	緯

序號	碼位	字形 (#1)	提交字形
177	7DD1	綠	綠
178	7DFC	緼	緼
179	7E4E	縶	縶
180	7E85	縶	縶
181	7F57	罗	罗
182	7F59	采	采
183	7F99	美	美
184	7FF6	翱	翱
185	806B	睽	睽
186	80A4	肤	肤
187	831F	葦	葦
188	8370	荳	荳

序號	碼位	字形 (#1)	提交字形
189	8380	宸	宸
190	839F	蒼	蒼
191	8402	菘	菘
192	847B	蕙	蕙
193	84B3	蒔	蒔
194	84D5	蕙	蕙
195	8532	蕙	蕙
196	8858	銜	銜
197	885C	銜	銜
198	8A09	訊	訊
199	8AB1	諛	諛
200	8B71	譚	譚

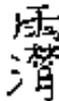
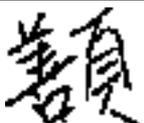
序號	碼位	字形 (#1)	提交字形
201	8CD0	賤	賤
202	8EC6	體	體
203	90DE	郎	郎
204	917F	釀	釀
205	9196	醞	醞
206	9268	鈦	鈦
207	9290	鋳	鋳
208	929E	鋳	鋳
209	935D	鋳	鋳
210	9372	鋳	鋳
211	93B6	鋳	鋳
212	945A	鑽	鑽

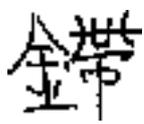
序號	碼位	字形 (#1)	提交字形
213	95D8	鬪	鬪
214	9629	阡	阡
215	96BA	雀	雀
216	96EB	雫	雫
217	970A	靈	靈
218	9836	頤	頤
219	986D	頰	頰
220	98DD	飛	飛
221	9A33	駙	駙
222	9AD9	高	高
223	9AEA	髮	髮
224	9CEF	鳳	鳳

序號	碼位	字形 (#1)	提交字形
225	9DE7	鸚	鸚
226	9EBA	麵	麵
227	9F8F	葬	葬
228	206D9	圳	圳
229	21429	叁	叁
230	21EBD	嶠	嶠
231	2256A	僑	僑
232	232A7	明	明
233	2340E	椀	椀
234	2352F	桶	桶
235	23D67	涓	涓
236	23F4E	淖	淖

序號	碼位	字形 (#1)	提交字形
237	24F88	暉	暉
238	24FAB	曜	曜
239	2590A	權	權
240	25DC0	簾	簾
241	25DDC	飶	飶
242	26B1E	芙	芙
243	26B73	莠	莠
244	26BB7	苜	苜
245	26C34	菰	菰
246	26C5C	菽	菽
247	26CF5	菡	菡
248	26E02	藻	藻

序號	碼位	字形 (#1)	提交字形
249	26FBA	莛	莛
250	275EA	衍	衍
251	279A4	誦	誦
252	279CB	誦	誦
253	27B02	讚	讚
254	27B4A	諫	諫
255	284AA	速	速
256	2866C	邨	邨
257	286EF	邨	邨
258	286F4	邨	邨
259	28A93	鎔	鎔
260	28F2A	隄	隄

序號	碼位	字形 (#1)	提交字形
261	291A4	霽	
262	29491	頤	
263	2954A	頤	
264	29561	頤	

序號	碼位	字形 (#1)	提交字形
265	29FC5	鳩	
266	2A04F	鵠	
267	2ACF5	橋 (#2)	
268	2B4BF	鐳 (#2)	

#1 Windows 7 的細明體

#2 Windows 10 的細明體