

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document**Title: Proposal for a revised Tangut character set for encoding in the SMP of the UCS****Source: Michael Everson, Nathan Hill, Guillaume Jacques, Andrew West, Viacheslav Zaytsev****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2009-04-08**

1. Introduction. This is a proposal to replace the set of Tangut characters under ballot on Amd.7 with an extended and consistently ordered set of characters. Extensive discussion with Tangutologists from China, France, Japan, Russia, the UK, and the USA has indicated that the current set of 5,910 characters does not meet the requirements of the user community. In particular Tangut users need to be able to represent all graphically distinct Tangut characters that are used in modern dictionaries and scholarly works, which the character repertoire under ballot does not allow. The set of 6,221 Tangut characters that we are proposing includes all graphically distinct characters used in Kyčanov 2006, Lǐ Fànwén 1997/2008 and Hán Xiǎománg 2004.

Previous documents on Tangut included:

N3297 *Proposal to encode Tangut characters in UCS Plane 1*. Richard Cook (UC Berkeley Script Encoding Initiative), 2007-05-09

N3307 *Tangut Background*. Richard Cook (UC Berkeley Script Encoding Initiative), 2007-09-01

N3338 *Response to UC Berkeley's proposals on Tangut*. China NB, 2007-09-16

N3343 *Expert feedback on Chinese NB input on WG2/N3297 Tangut Encoding Proposal*. Richard Cook, 2007-09-17

N3448 *Comments on N3297: Proposal to encode Tangut characters in UCS Plane 1 and Charts*. UK NB, 2008-04-19

N3467 *Comments on N3297: Proposal to encode Tangut characters in UCS Plane 1 and charts*. China and US NBs, 2008-04-22

N3495 *Proposal to encode Tangut Radicals and CJK Strokes in the UCS*. Michael Everson and Andrew West, 2008-09-01

N3496 *Review of Proposed Tangut Repertoire*. UK NB, 2008-09-01 (revised 2008-09-07)

N3498 *Expert Feedback on the proposed Tangut character set in PDAM 6.2*. Michael Everson and Andrew West, 2008-09-24

N3521 *The UCS Tangut Repertory*. Richard Cook and Ken Lunde, 2008-10-10

N3539 *Response from Tangut scholars of China on the Tangut Unicode proposal*. China NB, 2008-10-13

N3541 *Report from the Ad Hoc on Tangut*. Erkki I. Kolehmainen, 2008-10-13

N3577 *Proposal for a revised Tangut character set for encoding in the SMP of the UCS*. Michael Everson, Nathan Hill, Guillaume Jacques, Andrew West, Viacheslav Zaytsev, 2009-03-01

N3586 *Request for Tangut font and mappings from N3577 to Amendment 7 repertoire*. Deborah Anderson and Richard Cook, 2009-03-04

Table 1: WG2 Documents relating to Tangut

2. Character Repertoire. Extensive discussion with Tangutologists from China, France, Japan, Russia, the UK and the USA has indicated that the user community needs to be able to represent all graphically distinct Tangut characters that are used contrastively in modern dictionaries and scholarly works, even when they are variant character forms or erroneous character forms. Due to the complex structure of individual Tangut characters, and the fact that many characters are confusingly similar to other characters, Tangutologists are particularly concerned with the structural composition of characters, and need to be able to precisely represent the particular character under discussion. Thus the user community is not in favour of extensive unification of character variants. In addition, as Tangut is not a fully deciphered script, there may be disagreement amongst experts as to whether two similar character forms are simple glyph variants of the same character or whether they are semantically distinct characters. We believe that it is safest to err on the side of caution, and separately encode all characters that are used contrastively in contemporary and modern sources. This is in accord with the opinion expressed by Chinese Tangut experts in N3539.

The set of 6,221 Tangut characters proposed in this document includes all graphically distinct characters used in Kyčanov 2006, Lǐ Fànwén 1997/2008 and Hán Xiǎománg 2004. Although this does not cover all possible variant forms of Tangut characters that can be found in contemporary printed material, manuscripts and inscriptions, it does cover almost all of the characters that are used in modern Tangut scholarship, and we believe that it is a satisfactory basic set of Tangut characters. However, it is possible that further characters may need to be encoded at a future date.

3. Ordering Principles. It is essential to have a clearly defined system of character ordering for a large character set such as Tangut, both to facilitate the lookup of characters by users as well as to ensure that any additions to the character repertoire made during the encoding process can be correctly interpolated into the original set of characters. These ordering principles will be applied to the forthcoming revision of our Tangut radicals proposal, and may also be used for any future Tangut extension block.

Although almost all modern Tangut dictionaries and glossaries order characters by radical and stroke count, there is considerable variation in the set of radicals used as a basis for ordering, and so there is no universally accepted or standard character ordering system for Tangut. Most authors (e.g. Nishida, Sofronov, Shǐ Jīnbō, Lǐ Fànwén, and Hán Xiǎománg) choose the leftmost structural element of a character as its radical where possible, and the top, bottom or surrounding element if not; but some authors (e.g. Grinstead, Kolokolov and Kyčanov) choose the structural element at the bottom right of a character as its radical. Although both systems have their relative merits, we have decided to use a radical system based on the leftmost structural element as the basis for ordering the proposed character repertoire. The ordering principles described here are based on those used in Hán Xiǎománg 2004, with some modifications where necessary, and applied consistently to the entire set of characters.

The determination of radicals and the definition of stroke count and stroke ordering is necessarily a somewhat arbitrary task, but by following a clearly defined set of ordering principles we have aimed to produce a consistent and deterministic system of ordering for Tangut characters.

The following principles have been used to determine what radical a character is ordered under:

- the leftmost element or group of elements is chosen as the radical where possible (such as, for example, 彡 for 彡, and 彡 for 彡)
- where a character does not have an element or group of elements on the left side, the top, bottom or enclosing element is chosen as the radical (for example, 彡 for 彡, 彡 for 彡, and 彡 for 彡)
- where a character has both a top and bottom element (but no left side element), the top element is chosen as the radical (e.g. 彡 for 彡)
- characters that occur as a left side radical are ordered with the characters that use it as a radical, even if the character itself could be more naturally ordered under a different radical (for example, the

character 𠄎 is ordered together with the character 𠄎 under the radical 𠄎, rather than under the radical 𠄎)

- if the same radical shape occurs in different positions they are treated as separate radicals (e.g. 𠄎 and 𠄎 are treated as separate radicals, and 𠄎 and 𠄎 are treated as separate radicals)
- left side radicals and left-and-under radicals with the same basic glyph shape are treated as separate radicals (for example, 𠄎 and 𠄎 are treated as separate radicals, and 𠄎 and 𠄎 are treated as separate radicals).

Based on these principles we have identified a set of 527 radicals that can be used to order the proposed set of characters (note that these are only a subset of the radicals that are required to be encoded in order to cover different systems of radical classification). The relative ordering of these radicals, as well the ordering of characters within the same radical, is based on the nominal stroke count and stroke order of the character. We have identified eighteen basic stroke types used for Tangut characters, as shown in Table 2, each of which we assign a letter “A” through “R”.

Table 2: Tangut Strokes

Stroke ID	Typical Stroke	Example Radicals
A	一	二 卅 丰
B	丨	𠄎 丰
C	丿	𠄎 𠄎 𠄎
D	丶	𠄎 𠄎 𠄎 𠄎
E	㇇ ㇈	𠄎 𠄎
F	フ	𠄎 𠄎 𠄎
G	𠄎	𠄎
H	𠄎	𠄎
I	𠄎	𠄎
J	𠄎	𠄎 𠄎
K	𠄎	𠄎 𠄎 𠄎
L	𠄎	𠄎 𠄎
M	𠄎	𠄎 𠄎
N	𠄎	𠄎
O	𠄎	𠄎 𠄎
P	𠄎	𠄎 𠄎
Q	𠄎	𠄎 𠄎
R	𠄎	𠄎 𠄎

An alphabetic sort key is constructed for each radical from the letters corresponding to its constituent strokes. The order of letters in the sort key is based on the nominal order in which the strokes of the character should be written. Thus the radical 𠄎 is given a sort key of DCBOE based on the nominal writing order of its five constituent strokes. Note that although it is possible to reconstruct the stroke order of Tangut characters to a certain extent by analogy with the stroke order used for writing Han ideographs, some Tangut stroke constructions are unlike anything found in Chinese, and different scholars have different opinions as to what their correct stroke order should be, and even whether a stroke is a single stroke or two conjoined strokes. Although there may not be universal agreement on our

definition of stroke count and stroke order in all cases, we have attempted to be completely consistent, so that the same stroke construction always has the same alphabetic sort key wherever it occurs.

Radicals are ordered by their alphabetic sort key. Where two or more radicals have the same sort key the following principles are applied:

- left side radicals are ordered before top radicals with the same glyph shape (e.g. 𠄎 DCFABBB is ordered before 𠄎 DCFABBB)
- top radicals are ordered before bottom radicals with the same glyph shape (e.g. 𠄎 ABBB is ordered before 𠄎 ABBB)
- radicals with a final long slanting stroke are ordered immediately after the similar radical with a short slanting slope (e.g. 𠄎 CCCQ is ordered immediately after 𠄎 CCCQ)
- radicals with a stroke bending to the right are ordered immediately after the similar radical with a stroke that does not bend to the right (e.g. 𠄎 DCBE is ordered immediately after 𠄎 DCBE)
- radicals with crossing strokes are ordered after characters with non-crossing strokes (e.g. 𠄎 EAAAB is ordered after 𠄎 EAAAB, and 𠄎 EABEAA is ordered after 𠄎 EABEAA which is ordered after 𠄎 EABEAA)

Using these principles we order the 527 radicals as shown in Table 3.

Table 3 : Tangut Radicals used for Character Ordering

Number	Radical	Strokes	Stroke Order
1	一	1	A
2	丨	1	B
3	丿	1	C
4	㇇	1	F
5	丶	1	O
6	二	2	AA
7	𠃉	2	AB
8	𠃊	2	BB
9	𠃋	2	BE
10	𠃌	2	BE
11	𠃍	2	CB
12	𠃎	2	DA
13	𠃏	2	DC
14	𠃐	2	EA
15	𠃑	2	HO
16	𠃒	2	JC
17	𠃓	2	NB
18	𠃔	2	OF
19	𠃕	2	OG
20	𠃖	3	AAB
21	𠃗	3	AAB
22	𠃘	3	AAM

Number	Radical	Strokes	Stroke Order
23	㇇	3	ABA
24	㇈	3	ABB
25	㇉	3	ABB
26	㇊	3	ABE
27	㇋	3	ABO
28	㇌	3	ABR
29	㇍	3	AJC
30	㇎	3	BAA
31	㇏	3	BAE
32	㇐	3	BEB
33	㇑	3	CBB
34	㇒	3	CCC
35	㇓	3	CCQ
36	㇔	3	CCQ
37	㇕	3	CCQ
38	㇖	3	DAB
39	㇗	3	DAB
40	㇘	3	DCA
41	㇙	3	DCB
42	㇚	3	DCB
43	㇛	3	DDC
44	㇜	3	EAA
45	㇝	3	EAG
46	㇞	3	FBA
47	㇟	3	FBB
48	㇠	3	FCB
49	㇡	3	FCQ
50	㇢	3	HAB
51	㇣	3	HBB
52	㇤	3	KDD
53	㇥	3	LBO
54	㇦	3	OFB
55	㇧	4	AAAB
56	㇨	4	AABB
57	㇩	4	AABB
58	㇪	4	AABE
59	㇫	4	AAJC
60	㇬	4	ABBB
61	㇭	4	ABBB

Number	Radical	Strokes	Stroke Order
62	𠄎	4	ABEA
63	𠄏	4	ABEB
64	𠄐	4	ABFQ
65	𠄑	4	ACCQ
66	𠄒	4	AFCQ
67	𠄓	4	AKDD
68	𠄔	4	BAAE
69	𠄕	4	BAEA
70	𠄖	4	BBAE
71	𠄗	4	BEAA
72	𠄘	4	BECC
73	𠄙	4	CBAA
74	𠄚	4	CCBB
75	𠄛	4	CCBB
76	𠄜	4	CCBE
77	𠄝	4	CCCQ
78	𠄞	4	CCCQ
79	𠄟	4	CCCQ
80	𠄠	4	CCCQ
81	𠄡	4	CCQB
82	𠄢	4	CCQB
83	𠄣	4	CCQD
84	𠄤	4	CCQO
85	𠄥	4	CMCD
86	𠄦	4	CQBB
87	𠄧	4	DABE
88	𠄨	4	DABE
89	𠄩	4	DAHO
90	𠄪	4	DAJC
91	𠄫	4	DAOF
92	𠄬	4	DCAB
93	𠄭	4	DCBB
94	𠄮	4	DCBE
95	𠄯	4	DCBE
96	𠄰	4	DCBO
97	𠄱	4	DCCB
98	𠄲	4	DCCQ
99	𠄳	4	DCEA
100	𠄴	4	DCJC

Number	Radical	Strokes	Stroke Order
101	𠂇	4	DCJC
102	𠂈	4	EAAC
103	𠂉	4	EABC
104	𠂊	4	EABE
105	𠂋	4	EABE
106	𠂌	4	EACQ
107	𠂍	4	EAEA
108	𠂎	4	EAJC
109	𠂏	4	FAAB
110	𠂐	4	FABB
111	𠂑	4	FBBA
112	𠂒	4	FBBB
113	𠂓	4	FBBB
114	𠂔	4	FBCQ
115	𠂕	4	FBOE
116	𠂖	4	FCCQ
117	𠂗	4	HAAB
118	𠂘	4	HBBB
119	𠂙	4	JCBE
120	𠂚	4	KBOE
121	𠂛	4	KDBO
122	𠂜	4	LBOE
123	𠂝	4	OCCQ
124	𠂞	5	AABBB
125	𠂟	5	AABEA
126	𠂠	5	AABEB
127	𠂡	5	AABEB
128	𠂢	5	AACCQ
129	𠂣	5	ABAAB
130	𠂤	5	ABAAB
131	𠂥	5	ABBBA
132	𠂦	5	ABBBB
133	𠂧	5	ABEAA
134	𠂨	5	ABEAA
135	𠂩	5	ABEAA
136	𠂪	5	ABEAA
137	𠂫	5	ABEAA
138	𠂬	5	ABEAA
139	𠂭	5	ABFFQ

Number	Radical	Strokes	Stroke Order
140	𠄎	5	ACCCQ
141	𠄏	5	ACCQB
142	𠄐	5	ADCJC
143	𠄑	5	AEABE
144	𠄒	5	AFBBB
145	𠄓	5	AOCCQ
146	𠄔	5	BAAAB
147	𠄕	5	BAAAB
148	𠄖	5	BAAAB
149	𠄗	5	BAEAA
150	𠄘	5	BAFFQ
151	𠄙	5	BEAAA
152	𠄚	5	BECCQ
153	𠄛	5	CAAAB
154	𠄜	5	CAABB
155	𠄝	5	CBAAA
156	𠄞	5	CCBBB
157	𠄟	5	CCBBB
158	𠄠	5	CCBEA
159	𠄡	5	CCCCQ
160	𠄢	5	CCCQB
161	𠄣	5	CCCQB
162	𠄤	5	CCCQD
163	𠄥	5	CCCQD
164	𠄦	5	CCQCQ
165	𠄧	5	CCQOB
166	𠄨	5	CQCCQ
167	𠄩	5	DAABE
168	𠄪	5	DAAJC
169	𠄫	5	DABEA
170	𠄬	5	DACCQ
171	𠄭	5	DCAAB
172	𠄮	5	DCABB
173	𠄯	5	DCABB
174	𠄰	5	DCABE
175	𠄱	5	DCABO
176	𠄲	5	DCACQ
177	𠄳	5	DCAJC
178	𠄴	5	DCBAA

Number	Radical	Strokes	Stroke Order
179	𠄎	5	DCBEA
180	𠄏	5	DCBJC
181	𠄐	5	DCBOE
182	𠄑	5	DCBOP
183	𠄒	5	DCCCQ
184	𠄓	5	DCFBB
185	𠄔	5	DCFCQ
186	𠄕	5	DCKBB
187	𠄖	5	EAAAB
188	𠄗	5	EAAAB
189	𠄘	5	EAABE
190	𠄙	5	EAABE
191	𠄚	5	EAACQ
192	𠄛	5	EAAJC
193	𠄜	5	EABAE
194	𠄝	5	EABEA
195	𠄞	5	EABEB
196	𠄟	5	EACCQ
197	𠄠	5	EACCQ
198	𠄡	5	FAAAB
199	𠄢	5	FAABB
200	𠄣	5	FABBA
201	𠄤	5	FABBA
202	𠄥	5	FABBB
203	𠄦	5	FABBC
204	𠄧	5	FBAAA
205	𠄨	5	FBAAB
206	𠄩	5	FBAAB
207	𠄪	5	FBABB
208	𠄫	5	FBCCQ
209	𠄬	5	FCBEA
210	𠄭	5	HBAAB
211	𠄮	5	HBAAB
212	𠄯	5	HBCCQ
213	𠄰	5	JCCCQ
214	𠄱	5	KDAAB
215	𠄲	5	KDBBB
216	𠄳	5	KDBOE
217	𠄴	5	KDCCQ

Number	Radical	Strokes	Stroke Order
218	𠂇	5	OCCBB
219	冊	6	AABBBB
220	𠂇	6	AABEAA
221	𠂇	6	AACCCQ
222	𠂇	6	AADCAB
223	𠂇	6	AADCJC
224	𠂇	6	ABAAAB
225	𠂇	6	ABAEAA
226	𠂇	6	ABCCQB
227	𠂇	6	ABEAAA
228	𠂇	6	ACCCQB
229	𠂇	6	ACCQCQ
230	𠂇	6	AEABEA
231	𠂇	6	AKDBBB
232	𠂇	6	BAAAAB
233	𠂇	6	BAAEAA
234	𠂇	6	BAEAAA
235	𠂇	6	BAFBBB
236	𠂇	6	BEABBB
237	𠂇	6	BECCCQ
238	𠂇	6	BECCCQ
239	𠂇	6	CAABEB
240	𠂇	6	CBCCCQ
241	𠂇	6	CCBEAA
242	𠂇	6	CCBEAA
243	𠂇	6	CCBFFQ
244	𠂇	6	CCCQCQ
245	𠂇	6	CCQBEA
246	𠂇	6	CCQCCQ
247	𠂇	6	CCQCQB
248	𠂇	6	DAABEA
249	𠂇	6	DAACCQ
250	𠂇	6	DAAEAA
251	𠂇	6	DABBAA
252	𠂇	6	DABCCQ
253	𠂇	6	DABEAA
254	𠂇	6	DABEAA
255	𠂇	6	DABFCQ
256	𠂇	6	DACCCQ

Number	Radical	Strokes	Stroke Order
257	𠂇	6	DACCQO
258	𠂈	6	DCAAJC
259	𠂉	6	DCABEA
260	𠂊	6	DCABEB
261	𠂋	6	DCACCQ
262	𠂌	6	DCBAAA
263	𠂍	6	DCBAAB
264	𠂎	6	DCBABE
265	𠂏	6	DCBABE
266	𠂐	6	DCBBEA
267	𠂑	6	DCBCCQ
268	𠂒	6	DCBCCQ
269	𠂓	6	DCBEAA
270	𠂔	6	DCBEAA
271	𠂕	6	DCBECC
272	𠂖	6	DCCCCQ
273	𠂗	6	DCCQCQ
274	𠂘	6	DCEABE
275	𠂙	6	DCFABB
276	𠂚	6	DCFBEA
277	𠂛	6	DCFCCQ
278	𠂜	6	DCFQQC
279	𠂝	6	EAAAAB
280	𠂞	6	EAABBB
281	𠂟	6	EAABBE
282	𠂠	6	EAACCQ
283	𠂡	6	EABABE
284	𠂢	6	EABCCQ
285	𠂣	6	EABEAA
286	𠂤	6	EABEAA
287	𠂥	6	EABEAA
288	𠂦	6	EABECC
289	𠂧	6	EACCCQ
290	𠂨	6	EACCCQ
291	𠂩	6	EACCQD
292	𠂪	6	EAEABE
293	𠂫	6	FABBBA
294	𠂬	6	FABBBB
295	𠂭	6	FBAAAB

Number	Radical	Strokes	Stroke Order
296	𦉳	6	FBAAJC
297	𦉴	6	FBEAAA
298	𦉵	6	HBAAAB
299	𦉶	6	KBOCCQ
300	𦉷	6	KBOEAA
301	𦉸	6	KDDBBE
302	𦉹	6	LBOEAA
303	𦉺	6	OCAAAB
304	𦉻	6	OFBAAB
305	𦉼	7	AABEAAA
306	𦉽	7	AADCBEA
307	𦉽	7	AAFBEAA
308	𦉾	7	ABBAABE
309	𦉿	7	ABEACCQ
310	𦊀	7	ABECCCQ
311	𦊁	7	ABFQBEA
312	𦊂	7	ACCCQCB
313	𦊃	7	ACCQAAB
314	𦊄	7	ACCQCCQ
315	𦊅	7	ACCQCQB
316	𦊆	7	ACCQFCQ
317	𦊇	7	ADCACCQ
318	𦊈	7	ADCBEAA
319	𦊉	7	AKDABBB
320	𦊊	7	AKDABEB
321	𦊋	7	BAEABEA
322	𦊌	7	CCQCCCQ
323	𦊍	7	CCQCCQB
324	𦊎	7	CCQCQBE
325	𦊏	7	CCQDCCQ
326	𦊐	7	DABAEAA
327	𦊑	7	DABCCCQ
328	𦊒	7	DABFCCQ
329	𦊓	7	DACCCQB
330	𦊔	7	DCAABEB
331	𦊕	7	DCABAAB
332	𦊖	7	DCABABE
333	𦊗	7	DCABABE
334	𦊘	7	DCABCCQ

Number	Radical	Strokes	Stroke Order
335	𠂇	7	DCABEAA
336	𠂇	7	DCABEAA
337	𠂇	7	DCACQCQ
338	𠂇	7	DCAMAAA
339	𠂇	7	DCBAAAB
340	𠂇	7	DCBAAJC
341	𠂇	7	DCBBEAA
342	𠂇	7	DCBCCCQ
343	𠂇	7	DCBOCCQ
344	𠂇	7	DCBOEAA
345	𠂇	7	DCCCQCQ
346	𠂇	7	DCDAABE
347	𠂇	7	DCDCCCQ
348	𠂇	7	DCEAABE
349	𠂇	7	DCEAACQ
350	𠂇	7	DCEABEA
351	𠂇	7	DCEACCQ
352	𠂇	7	DCFAAAB
353	𠂇	7	DCFABBB
354	𠂇	7	DCFABBB
355	𠂇	7	DCFABBB
356	𠂇	7	DCFBAAB
357	𠂇	7	DCFBEAA
358	𠂇	7	DCKDAAB
359	𠂇	7	DCKDBBB
360	𠂇	7	EAABBAAE
361	𠂇	7	EAABEAA
362	𠂇	7	EAABEAA
363	𠂇	7	EAACCCQ
364	𠂇	7	EACBAAB
365	𠂇	7	EACCQCQ
366	𠂇	7	EADCAAB
367	𠂇	7	EADCBOE
368	𠂇	7	EAEABEA
369	𠂇	7	EAEABEB
370	𠂇	7	EAMAAAA
371	𠂇	7	FABBAAB
372	𠂇	7	FABBABE
373	𠂇	7	FABBBBE

Number	Radical	Strokes	Stroke Order
374	爻	7	FABBCCQ
375	爻	7	FABCCCQ
376	彳	7	FBAAAAB
377	𠂔	7	FBAABMC
378	𠂔	7	KDAABBB
379	彳	7	KDBOEAA
380	𠂔	7	KDDBCCQ
381	爻	7	KDDCCCQ
382	𠂔	7	OCCBFFQ
383	𠂔	8	AABBBBAE
384	爻	8	AABBCCQ
385	爻	8	AACCQCCQ
386	𠂔	8	ABBABEAA
387	𠂔	8	ABBCCCQ
388	𠂔	8	ABBDCBB
389	𠂔	8	ABBDCJC
390	𠂔	8	ABBFAAB
391	𠂔	8	ABBBLBOE
392	𠂔	8	ABDCFCCQ
393	𠂔	8	ABEAAABE
394	𠂔	8	ABEA AFCQ
395	𠂔	8	ABEABEAA
396	𠂔	8	ABECCQCCQ
397	𠂔	8	ABFQABBB
398	𠂔	8	ABFQABFQ
399	𠂔	8	ABFQABFQ
400	爻	8	ACCQCCQ
401	𠂔	8	BBAABEAA
402	𠂔	8	BEAACCCQ
403	𠂔	8	CCCQBEAA
404	𠂔	8	DAABBBEA
405	𠂔	8	DABAAABE
406	𠂔	8	DABDCKBB
407	𠂔	8	DADCFCCQ
408	𠂔	8	DCAABEAA
409	𠂔	8	DCAACCQB
410	𠂔	8	DCABAABB
411	𠂔	8	DCABABEM
412	𠂔	8	DCABACCQ

Number	Radical	Strokes	Stroke Order
413	𠂇	8	DCAB AFCQ
414	𠂈	8	DCAB EAAA
415	𠂉	8	DCAC BAAA
416	𠂊	8	DCAC CB EA
417	𠂋	8	DCAC CQCQ
418	𠂌	8	DCAD CBOE
419	𠂍	8	DCAD CFCQ
420	𠂎	8	DCBA BEAA
421	𠂏	8	DCBC CQCQ
422	𠂐	8	DCBE CCCQ
423	𠂑	8	DCDA BEAA
424	𠂒	8	DCDC EABE
425	𠂓	8	DCEA ABEA
426	𠂔	8	DCEA BCCQ
427	𠂕	8	DCEA BEAA
428	𠂖	8	DCEA BECC
429	𠂗	8	DCEA CCCQ
430	𠂘	8	DCEA DCBE
431	𠂙	8	DCFC CCCQ
432	𠂚	8	EAAC CCQB
433	𠂛	8	EAAE AAAB
434	𠂜	8	EAAFB EAA
435	𠂝	8	EAAK DBBB
436	𠂞	8	EADCA CCQ
437	𠂟	8	FABB ACCQ
438	𠂠	8	FABB AJC
439	𠂡	8	FABB EAA
440	𠂢	8	FABB CB EA
441	𠂣	8	FABB CCQO
442	𠂤	8	FABB DCBB
443	𠂥	8	FABB EABE
444	𠂦	8	FBAAB EAC
445	𠂧	8	FBBB CCCQ
446	𠂨	8	KDDCC QBE
447	𠂩	8	LBOE ABBB
448	𠂪	9	AABBB ACCQ
449	𠂫	9	AACCC QCCQ
450	𠂬	9	ABBBB EAA
451	𠂭	9	ABBBCC CQB

Number	Radical	Strokes	Stroke Order
452	莽	9	ABBBCQBEA
453	莽	9	ABBBKDBBB
454	夏	9	ABEAACCCQ
455	𦰇	9	BAEAAABBB
456	𦰇	9	CCCQABBBA
457	𦰇	9	DABDCFCCQ
458	𦰇	9	DCAABEAAA
459	𦰇	9	DCABABEAA
460	𦰇	9	DCABABEAA
461	𦰇	9	DCADCBCCQ
462	𦰇	9	DCBCCCQCQ
463	𦰇	9	DCBOEABBB
464	𦰇	9	DCFABBBBA
465	𦰇	9	DCFABBCCQ
466	𦰇	9	DCFABBCCQ
467	𦰇	9	DCFQQCCCQ
468	𦰇	9	EAAJCDABE
469	𦰇	9	EAAKDBBBB
470	𦰇	9	EABECCCQA
471	𦰇	9	EABECCCQA
472	𦰇	9	EACBECCCQ
473	𦰇	9	EACCQCCCQ
474	𦰇	9	FABBABEAA
475	𦰇	9	FABBBACCQ
476	𦰇	9	FABBCBEAA
477	𦰇	9	FABBEACCQ
478	𦰇	9	FABBFABBA
479	𦰇	9	FCCQABBBA
480	𦰇	10	AABBBCQCQ
481	𦰇	10	AABBBKDBBB
482	𦰇	10	AACCCQDCAB
483	𦰇	10	ABBBACCCQB
484	𦰇	10	ABBBFBAAAB
485	𦰇	10	ABEAAABBBA
486	𦰇	10	ABEDCABCCQ
487	𦰇	10	DABEAACCCQ
488	𦰇	10	DCAABBBBEA
489	𦰇	10	DCBABECCCQ
490	𦰇	10	DCEABECCCQ

Number	Radical	Strokes	Stroke Order
491	𦰇	10	DCFABBBAAAB
492	𦰈	10	DCFABBBCCQ
493	𦰉	10	DCFCQABBBA
494	𦰊	10	EACCQABBBA
495	𦰋	10	EACFABCCCQ
496	𦰌	10	FABBAABCCQ
497	𦰍	10	FABBCCQCCQ
498	𦰎	10	FABBDCBAAB
499	𦰏	10	FABBDCFCCQ
500	𦰐	11	AABBBAABBBA
501	𦰑	11	AABBBBAECCQ
502	𦰒	11	AABBBCCQCCQ
503	𦰓	11	AABBBDCFCCQ
504	𦰔	11	AABBBECCCQA
505	𦰕	11	AABBBECCCQD
506	𦰖	11	ABBBAABCCBB
507	𦰗	11	ABBBAABCCBB
508	𦰘	11	ABEAADCBCCCQ
509	𦰙	11	ABFQBEAABBB
510	𦰚	11	DACCCQBCCQ
511	𦰛	11	DCACCQABBBA
512	𦰜	11	EAAJCKDDBBE
513	𦰝	11	FABBBBECCCQ
514	𦰞	11	KDDBAEDACCQ
515	𦰟	12	ABECCCQCCQ
516	𦰠	12	DABABBBKDBBB
517	𦰡	12	DCFABBBBEAAA
518	𦰢	12	DCFABBBBEAAA
519	𦰣	12	DCFABBBDCBOE
520	𦰤	12	EAACCCQBCCQ
521	𦰥	12	FABBDCBBEABE
522	𦰦	13	ABBBDCBBEAAAB
523	𦰧	13	CCCQABEAACCCQ
524	𦰨	13	DCFABBBCCQABE
525	𦰩	13	DCFABBBCCQABE
526	𦰪	13	DCFABBBOCAAAB
527	𦰫	16	DCFABBBCCQAABCCQ

The characters assigned to each radical are ordered by residual stroke count and stroke order using the same principles used for ordering radicals. The resultant ordering of the proposed set of 6,221 characters is shown in Appendix A. Appendix A is available in a separate document, N3577R-A.

Note that the ideographic description sequences in Appendix A show the minimal possible decomposition of characters into their component parts using an extension to the set of radicals proposed in N3495 (Proposal to encode Tangut Radicals and CJK Strokes in the UCS). By applying the IDS decompositions of this set of radicals it would be possible to generate maximally decomposed IDS sequences for the proposed set of Tangut characters.

4. Source Mappings. Appendix B shows the mappings of the proposed set of characters to Kyčanov 2006, Lǐ Fàn wén 1997/2008 and Hán Xiǎo máng 2004, as well as to the corresponding character or characters in PDAM7 (because of unifications in the original proposal there is a one-to-many relationship between the PDAM7 Tangut repertoire and our proposed repertoire). Appendix B is available in a separate document, N3577R-B. In the Source Mapping Table, the following conventions are used:

- Col.1 = Proposed Code Point
- Col.2 = Hán Xiǎo máng 2004
- Col.3 = Lǐ Fàn wén 1997
- Col.4 = Lǐ Fàn wén 2008
- Col.5 = Kyčanov 2006
- Col.6 = PDAM7 Code Point (there are many-to-one mappings from the PDAM7 code points to the new code points because of unifications in the PDAM7 repertoire)
- Col.7 = Notes

5. Bibliography

- Grinstead, Eric. 1972. *Analysis of the Tangut Script* (Scandinavian Institute of Asian Studies Monograph Series No.10).
- Hán Xiǎo máng (韓小忙). 2004. 西夏文正字研究 (= *Xīxiàwén Zhèngzì Yánjiū*).
- Lǐ Fàn wén (李範文). 2008. 夏漢字典 (= *Xià-Hàn Zìdiàn*). Beijing.
- Lǐ Fàn wén (李範文). 1997. 夏漢字典 (= *Xià-Hàn Zìdiàn*). Beijing.
- Lǐ Fàn wén (李範文). 1986. 同音研究 (= *Tóngyīn Yánjiū*). Yinchuan.
- Keping, K. B. (К. Б. Кепинг) et al. 1969. *Море писъмен* (= *More pis'men*) [The Sea of Characters]. Moscow.
- Kolokolov, V. S. (В. С. Колоколов) and E. I. Kyčanov (Е. И. Кычанов). 1966. *Китайская классика в тангутском переводе* (= *Kitajskaja klassika v tangutskom perevode*) [Chinese Classics in Tangut Translation]. Moscow.
- Kyčanov, E. I. (Е. И. Кычанов). 2006. *Словарь тангутского (Си Ся) языка* (= *Slovar' tangutskogo (Si Sja) jazyka*) [Tangut-Russian-English-Chinese Dictionary]. St. Petersburg and Kyoto.
- Nakajima Motoki (中嶋幹起) et al 2000. 電腦処理《文海宝韻》研究 (= *Dennō shori 'Bunkai Hōin' kenkyū*). Tokyo.
- Nevskij, N. A. (Н. А. Невский). 1960. *Тангутская филология: Исследования и словарь* (= *Tangut-skaja filologija: Issledovanija i slovar'*) [Tangut philology: Researches and dictionary]. Moscow.
- Nishida Tatsuo (西田龍雄). 1966. 西夏文小字典 (= *Seikabun Shōjiten*). In 西夏語の研究 (= *Seikago no kenkyū*) [A Study of the Hsi-Hsia Language] (1964-1966) vol.2. Tokyo.
- Shǐ Jīnbō (史金波) et al. 1983. 文海研究 (= *Wénhǎi Yánjiū*). Beijing.
- Sofronov M. V. (М. В. Софронов). 1968. *Грамматика тангутского языка* (= *Grammatika tangut-skogo jazyka*). Moscow.

A. Administrative

1. Title

Proposal for a revised Tangut character set for encoding in the SMP of the UCS

2. Requester's name

Michael Everson, Nathan Hill, Guillaume Jacques, Andrew West, Viacheslav Zaytsev

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2009-04-08

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Tangut Radicals.

1c. The proposal is for addition of character(s) to an existing block

1d. Name of the existing block

2. Number of characters in proposal

6,221.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category F.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Co-authors are Tangutologists

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Scholars.

4a. The context of use for the proposed characters (type of use; common or rare)

Rare.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In scholarly publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?