

ISO/IEC JTC1/SC2/WG2
L2/04-221
2004-06-07

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Document Type: Working Group Document

Title: Proposal to add a block of CJK Unified Basic Strokes to the UCS

Source: Thomas Bishop and Richard Cook

Status: Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2004-06-07

This proposal is a continuation of work presented previously in L2/03-387, L2/03-404, and L2/03-420. These three earlier documents provide background information along with a complete specification for CDL, the Chinese Character Description Language [a.k.a. C(UIG)DL the CJK Unified Ideographic Glyph Description Language].



























In particular, L2/03-420 presents the set of basic stroke types and subtypes, along with distributional and distinctive feature analyses. L2/03-420 makes explicit that the set of CDL Basic Strokes is based squarely upon the representative Song-style CJK forms appearing in UCS code charts. Typography and stroke counting presume stroke type identification, and CDL is reliant upon modern PRC calligraphic norms and upon the primary Unihan.txt lexical sources.



























At present the CDL database includes complete descriptions for more than 56,000 CJK Unified Ideographs. In the course of creating these descriptions several very rare or somewhat ill-defined stroke types have been found which the current CDL only approximates. For this reason, in addition to the set of 39 characters in the current proposal, it is recommended that the overall block size be set at 64, reserving 25 character positions for future refinement of the system.














The CJK Unified Basic Strokes presented here may be divided into classes according to encoding status. L2/03-420 presents this classification in terms of UCS vs. PUA mappings. The problem of PUA mappings for basic stroke types, emphasizing the special nature of this set of forms, has also presented itself with regard to HKSCS mappings. CDL mappings to HKSCS PUA forms have been prepared, and if the current proposal is approved it might prove useful to employ these mappings.

The encoding of this block of CJK Unified Basic Strokes is justified in terms of the special characteristics of these script entities. Their semantics is highly specialized, as they are used for particular indexing purposes. They are not all equally common as independent characters. CDL defines each type on a traditional basis in terms of a specific number of control points, specific segments, and specific behaviors.

UCS names proposed in the chart below derive from those appearing in L2/03-420.

#	USV	Form	Unified	Name
1	U+XX00			CJK UNIFIED BASIC STROKE H
2	U+XX01			CJK UNIFIED BASIC STROKE T
3	U+XX02			CJK UNIFIED BASIC STROKE S
4	U+XX03			CJK UNIFIED BASIC STROKE SG
5	U+XX04			CJK UNIFIED BASIC STROKE P
6	U+XX05			CJK UNIFIED BASIC STROKE WP
7	U+XX06			CJK UNIFIED BASIC STROKE SP
8	U+XX07			CJK UNIFIED BASIC STROKE D
9	U+XX08			CJK UNIFIED BASIC STROKE N
10	U+XX09			CJK UNIFIED BASIC STROKE DN
11	U+XX0A			CJK UNIFIED BASIC STROKE PN
12	U+XX0B			CJK UNIFIED BASIC STROKE TN
13	U+XX0C			CJK UNIFIED BASIC STROKE TPN

#	USV	Form	Unified	Name
14	U+XX0D			CJK UNIFIED BASIC STROKE HZ
15	U+XX0E			CJK UNIFIED BASIC STROKE HP
16	U+XX0F			CJK UNIFIED BASIC STROKE HG
17	U+XX10			CJK UNIFIED BASIC STROKE SZ
18	U+XX11			CJK UNIFIED BASIC STROKE SW
19	U+XX12			CJK UNIFIED BASIC STROKE ST
20	U+XX13			CJK UNIFIED BASIC STROKE PZ
21	U+XX14			CJK UNIFIED BASIC STROKE PD
22	U+XX15			CJK UNIFIED BASIC STROKE PG
23	U+XX16			CJK UNIFIED BASIC STROKE WG
24	U+XX17			CJK UNIFIED BASIC STROKE XG
25	U+XX18			CJK UNIFIED BASIC STROKE HZZ
26	U+XX19			CJK UNIFIED BASIC STROKE HZW

#	USV	Form	Unified	Name
27	U+XX1A		↓	CJK UNIFIED BASIC STROKE HZT
28	U+XX1B		𠂇 𠂈 𠂉	CJK UNIFIED BASIC STROKE HZG
29	U+XX1C		𠂊	CJK UNIFIED BASIC STROKE HXG
30	U+XX1D		𠂋	CJK UNIFIED BASIC STROKE SZZ
31	U+XX1E		𠂌 𠂍	CJK UNIFIED BASIC STROKE SZP
32	U+XX1F		𠂎	CJK UNIFIED BASIC STROKE SWG
33	U+XX20		𠂏	CJK UNIFIED BASIC STROKE HZZZ
34	U+XX21		𠂐	CJK UNIFIED BASIC STROKE HZZP
35	U+XX22		乙 𠂑	CJK UNIFIED BASIC STROKE HZWG
36	U+XX23		𠂒	CJK UNIFIED BASIC STROKE HPWG
37	U+XX24		𠂓	CJK UNIFIED BASIC STROKE SZZG
38	U+XX25		𠂔	CJK UNIFIED BASIC STROKE HZZZG
39	U+XX26		○	CJK UNIFIED BASIC STROKE O

#	USV	Form	Unified	Name
40	U+XX27			(This position shall not be used.)
41	U+XX28			(This position shall not be used.)
42	U+XX29			(This position shall not be used.)
43	U+XX2A			(This position shall not be used.)
44	U+XX2B			(This position shall not be used.)
45	U+XX2C			(This position shall not be used.)
46	U+XX2D			(This position shall not be used.)
47	U+XX2E			(This position shall not be used.)
48	U+XX2F			(This position shall not be used.)
49	U+XX30			(This position shall not be used.)
50	U+XX31			(This position shall not be used.)
51	U+XX32			(This position shall not be used.)
52	U+XX33			(This position shall not be used.)

#	USV	Form	Unified	Name
53	U+XX34			(This position shall not be used.)
54	U+XX35			(This position shall not be used.)
55	U+XX36			(This position shall not be used.)
56	U+XX37			(This position shall not be used.)
57	U+XX38			(This position shall not be used.)
58	U+XX39			(This position shall not be used.)
59	U+XX3A			(This position shall not be used.)
60	U+XX3B			(This position shall not be used.)
61	U+XX3C			(This position shall not be used.)
62	U+XX3D			(This position shall not be used.)
63	U+XX3E			(This position shall not be used.)
64	U+XX3F			(This position shall not be used.)

A. Administrative

1. Title:

Proposal to add a block of CJK Unified Basic Strokes to the UCS.

2. Requesters:

Tom Bishop <wenlin@wenlin.com> and Richard Cook <rscook@unicode.org>

3. Requester type:

Individual contribution

4. Submission date:

June 7, 2004

5. Requester's reference (if applicable):

NA

6. Choose one of the following:

This is a complete proposal.

B. Technical - General

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

Yes.

1b. Proposed name of script:

A new block in Plane 1, to be named "CJK Unified Basic Strokes".

2. Number of characters in proposal:

39, total block size of 64.

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary

4a. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):

Level 1

4b. Is a rationale provided for the choice?

Yes.

4c. If Yes, reference:

Spacing Characters.

5. Is a repertoire including character names provided?

Yes.

5a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

5b. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Bishop and Cook

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Bishop and Cook, Wenlin and Fontographer

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes, see L2/03-420. HKSCS mappings are also available.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. Special encoding issues: 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)?

No.

9. Additional Information: 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.

See L2/03-420.

C. Technical - Justification

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

No.

2. Has contact been made to members of the user community:

Yes.

2a. If YES, with whom?

IRG. Documents in C.2b were presented at IRG #21

2b. If YES, available relevant documents:

L2/03-387, L2/03-404, L2/03-420.

3. Information on the user community for the proposed characters is included?

Yes.

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

These characters are commonly used in calligraphy manuals, typography, indexing.

4b. Reference

See C.2b above.

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

Yes.

5b. If YES, where?

CJK calligraphy, typography, indexing, and info-tech

5c. Reference

See C.2b above.

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

Perhaps not.

6a. If YES, is a rationale provided?

NA

6b. If YES, reference

NA

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

Yes, in a contiguous block.

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

No.

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

No.

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

Yes, some representative forms might be similar in appearance.

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

Yes.

10b. If YES, reference

Yes.

11. Does the proposal include use of combining characters and/or use of composite sequences?

No.

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

No.

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

No.