



ISO/IEC JTC1/SC2/WG2 N1685

DATE: 1998-01-18

ISO/IEC JTC1/SC2/WG2
 Universal Multiple-Octet Coded Character Set (UCS) - ISO/IEC 10646
 Secretariat: ANSI

DOC TYPE: Expert contribution
TITLE: Proposal to encode Brahmi in Plane 1 of ISO/IEC 10646-2
SOURCE: Michael Everson, EGT (IE)
PROJECT: JTC1.02.18.02
STATUS: Proposal.
ACTION ID: FYI
DUE DATE: --
DISTRIBUTION: Worldwide
MEDIUM: Paper and web
NO. OF PAGES: 3 (printed at 80%)

A. Administrative

1. Title	Proposal to encode Brahmi in Plane 1 of ISO/IEC 10646-2
2. Requester's name	Michael Everson
3. Requester type	Expert request
4. Submission date	1998-01-18
5. Requester's reference	
6a. Completion	This is a complete proposal.
6b. More information to be provided?	No

B. Technical -- General	
1a. New script? Name?	Yes. Brahmi
1b. Addition of characters to existing block? Name?	No.
2. Number of characters	58
3. Proposed category	Category C
4. Proposed level of implementation and rationale	Level 2
5a. Character names included in proposal?	Yes
5b. Character names in accordance with guidelines?	Yes
5c. Character shapes reviewable?	Yes
6a. Who will provide computerized font?	Michael Everson
6b. Font currently available?	Michael Everson
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes.
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	No
8. Does the proposal address other aspects of character data processing?	Yes

C. Technical -- Justification	
1. Contact with the user community?	No.
2. Information on the user community?	Brahmi enjoys mostly scholarly use.
3a. The context of use for the proposed characters?	Used to write Buddhist and other texts in northern India.
3b. Reference	See below.
4a. Proposed characters in current use?	Yes.
4b. Where?	By scholars.
5a. Characters should be encoded entirely in BMP?	No. Plane 1
5b. Rationale	Accordance with the Roadmap.
6. Should characters be kept in a continuous range?	Yes
7a. Can the characters be considered a presentation form of an existing character or character sequence?	No.
7b. Where?	
7c. Reference	
8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?	No
8b. Where?	
8c. Reference	
9a. Combining characters or use of composite sequences included?	Yes, the usual Indic vowel matras.
9b. List of composite sequences and their corresponding glyph images provided?	No, a comprehensive list for Annex B will be provided at PDAM time.
10. Characters with any special properties such as control function, etc. included?	No
D. SC2/WG2 Administrative	
To be completed by SC2/WG2	
1. Relevant SC 2/WG 2 document numbers:	
2. Status (list of meeting number and corresponding action or disposition)	
3. Additional contact to user communities, liaison organizations etc.	
4. Assigned category and assigned priority/time frame	
Other Comments	

The script known as Brahmi was derived around the 5th century BCE from one of the old North Semitic alphabets, though a reanalysis of the glyphs was made and new characters were invented. Brahmi is the ancestor of the modern scripts in India and Southeast Asia. The Brahmi default directionality is LTR, though in some of the oldest texts RTL is found. Addition of vowel signs to the consonants results in conjoined syllables, a bit different from the practice of most Brahmic scripts; a fairly large font repertoire

is required to render Brahmi.

Issues:

Are there numbers or punctuation?

Are the glyphs correct? Is there a glyph for an explicit virama?

Would there be any value in encoding Brahmi parallel to Devanagari and the other Indic scripts in ISO/IEC 10646? The relative order of characters here is the same (i.e., ANUSVARA precedes initial vowels, then consonants, then vowel signs and VIRAMA).

According to the relatively scant sample text evidence I have to hand, the only conjunct formed is with a consonant and a following -YA. However, the VIRAMA is used in this proposal in preference to a SUBJOINED YA. It seemed that it would be the height of absurdity not to use UCS Brahmic coding conventions to encode Brahmi. In any case there may be other conjuncts not present in the small text sample given in Haarmann.

U+0001xx00	BRAHMI SIGN ANUSVARA
U+0001xx01	BRAHMI LETTER A
U+0001xx02	BRAHMI LETTER AA
U+0001xx03	BRAHMI LETTER I
U+0001xx04	BRAHMI LETTER II
U+0001xx05	BRAHMI LETTER U
U+0001xx06	BRAHMI LETTER UU
U+0001xx07	BRAHMI LETTER E
U+0001xx08	BRAHMI LETTER AI
U+0001xx09	BRAHMI LETTER O
U+0001xx0A	BRAHMI LETTER AU
U+0001xx0B	BRAHMI LETTER VOCALIC R
U+0001xx0C	BRAHMI LETTER KA
U+0001xx0D	BRAHMI LETTER KHA
U+0001xx0E	BRAHMI LETTER GA
U+0001xx0F	BRAHMI LETTER GHA
U+0001xx00	BRAHMI LETTER NGA
U+0001xx01	BRAHMI LETTER CA
U+0001xx02	BRAHMI LETTER CHA
U+0001xx03	BRAHMI LETTER JA
U+0001xx04	BRAHMI LETTER JHA
U+0001xx05	BRAHMI LETTER NYA
U+0001xx06	BRAHMI LETTER TTA
U+0001xx07	BRAHMI LETTER TTHA
U+0001xx08	BRAHMI LETTER DDA
U+0001xx09	BRAHMI LETTER DDHA
U+0001xx0A	BRAHMI LETTER NNA
U+0001xx0B	BRAHMI LETTER TA
U+0001xx0C	BRAHMI LETTER THA
U+0001xx0D	BRAHMI LETTER DA
U+0001xx0E	BRAHMI LETTER DHA
U+0001xx0F	BRAHMI LETTER NA
U+0001xx00	BRAHMI LETTER PA
U+0001xx01	BRAHMI LETTER PHA
U+0001xx02	BRAHMI LETTER BA
U+0001xx03	BRAHMI LETTER BHA
U+0001xx04	BRAHMI LETTER MA
U+0001xx05	BRAHMI LETTER YA
U+0001xx06	BRAHMI LETTER RA
U+0001xx07	BRAHMI LETTER LA
U+0001xx08	BRAHMI LETTER LLA
U+0001xx09	BRAHMI LETTER VA
U+0001xx0A	BRAHMI LETTER SHA
U+0001xx0B	BRAHMI LETTER SZA
U+0001xx0C	BRAHMI LETTER SSA
U+0001xx0D	BRAHMI LETTER SA
U+0001xx0E	BRAHMI LETTER HA
U+0001xx0F	(This position shall not be used)
U+0001xx00	(This position shall not be used)
U+0001xx01	(This position shall not be used)
U+0001xx02	BRAHMI VOWEL SIGN AA
U+0001xx03	BRAHMI VOWEL SIGN I
U+0001xx04	BRAHMI VOWEL SIGN II
U+0001xx05	BRAHMI VOWEL SIGN U
U+0001xx06	BRAHMI VOWEL SIGN UU
U+0001xx07	BRAHMI VOWEL SIGN E
U+0001xx08	BRAHMI VOWEL SIGN AI
U+0001xx09	BRAHMI VOWEL SIGN O
U+0001xx0A	BRAHMI VOWEL SIGN AU
U+0001xx0B	BRAHMI VOWEL SIGN VOCALIC R
U+0001xx0C	BRAHMI SIGN VIRAMA
U+0001xx0D	(This position shall not be used)
U+0001xx0E	(This position shall not be used)

	xx0	xx1	xx2	xx3
0	◌̣	◌̥	◌̦	
1	◌̧	◌̨	◌̩	
2	◌̪	◌̫	◌̬	◌̭
3	◌̮	◌̯	◌̰	◌̱
4	◌̲	◌̳	◌̴	◌̵
5	◌̶	◌̷	◌̸	◌̹
6	◌̺	◌̻	◌̼	◌̽
7	◌̾	◌̿	◌̻	◌̼
8	◌̽	◌̾	◌̿	◌̻
9	◌̻	◌̼	◌̽	◌̾
A	◌̾	◌̿	◌̻	◌̼
B	◌̧	◌̨	◌̩	◌̪
C	+	◌̣	◌̥	◌̦
D	◌̧	◌̨	◌̩	
E	◌̪	◌̫	◌̬	
F	◌̮	◌̯		

U+0001xx0F (This position shall not be used)

Bibliography

Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8

Haarmann, Harald. 1990. *Universalgeschichte der Schrift*. Frankfurt/Main; New York: Campus. ISBN 3-593-34346-0

HTML Michael Everson, everson@indigo.ie, <http://www.indigo.ie/egt>, Dublin, 1998-01-18