

L2/98-031



ISO/IEC JTC1/SC2/WG2 N1684

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 Avestan in the BMP of ISO/IEC 10646
SOURCE: Michael Everson, EGT (IE)
PROJECT: JTC1.02.18.01
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 Avestan 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. Avestan
1b. Addition of characters to existing block? Name?	No.
2. Number of characters	61
3. Proposed category	Category B.1
4. Proposed level of implementation and rationale	Level 1
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?	Yes. Joseph Peterson, Jan Pieter Kunst.
2. Information on the user community?	Avestan enjoys both scholarly and ecclesiastical use.
3a. The context of use for the proposed characters?	Used to represent texts in the Avestan and Old Persian languages.
3b. Reference	See below.
4a. Proposed characters in current use?	Yes.
4b. Where?	By scholars and Zoroastrians.
5a. Characters should be encoded entirely in BMP?	Yes
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?	No.
9b. List of composite sequences and their corresponding glyph images provided?	No.
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:	<input type="checkbox"/>
2. Status (list of meeting number and corresponding action or disposition)	<input type="checkbox"/>
3. Additional contact to user communities, liaison organizations etc.	<input type="checkbox"/>
4. Assigned category and assigned priority/time frame	<input type="checkbox"/>
Other Comments	<input type="checkbox"/>

The script known as Avestan is related to the Arabic alphabet. It is a true superset of the consonantal alphabet Pahlavi, and it is proposed here to unify the two scripts (i.e. to subsume Pahlavi into Avestan). This proposal is similar to the proposal of Rick McGowan in UTR #3. The Avestan default directionality is RTL. Unlike Arabic, the numbers seem also to have RTL directionality (but see the issue on numbers

below).

Issues:

Are ligatures obligatory? Some ligatures are formed and Pahlavi and Avestan fonts will need to take those into account. If the ligatures are not obligatory, then ZWJ should be used to make them.

Faulmann gives this set of numbers. This needs to be looked into with more modern sources and experts.

Is the punctuation coded correctly?

The names given here are versions of their Latin transliterations. Do actual names exist for these characters?

Two forms of Y and V are coded here; these are found in a current Avestan font set and the two are included here along the same lines as Greek has SIGMA and FINAL SIGMA and Hebrew has PE and FINAL PE. Avestan, though it looks like Arabic, is much more strongly alphabetic, like Greek and Hebrew, and it would be better to follow those scripts as models than to force the character/glyph model onto this script. Again, existing fonts encode initial and medial Y and V as separate characters.

The hyphen may be unifiable.

U+0001xx00 AVESTAN LETTER A
 U+0001xx00 AVESTAN LETTER AA
 U+0001xx00 AVESTAN LETTER AE
 U+0001xx00 AVESTAN LETTER AEE
 U+0001xx00 AVESTAN LETTER E
 U+0001xx00 AVESTAN LETTER EE
 U+0001xx00 AVESTAN LETTER O
 U+0001xx00 AVESTAN LETTER OO
 U+0001xx00 AVESTAN LETTER AO
 U+0001xx00 AVESTAN LETTER AN
 U+0001xx00 AVESTAN LETTER I
 U+0001xx00 AVESTAN LETTER II
 U+0001xx00 AVESTAN LETTER U
 U+0001xx00 AVESTAN LETTER UU
 U+0001xx00 AVESTAN LETTER K
 U+0001xx00 AVESTAN LETTER G
 U+0001xx00 AVESTAN LETTER GH
 U+0001xx00 AVESTAN LETTER X
 U+0001xx00 AVESTAN LETTER C
 U+0001xx00 AVESTAN LETTER J
 U+0001xx00 AVESTAN LETTER T
 U+0001xx00 AVESTAN LETTER D
 U+0001xx00 AVESTAN LETTER DH
 U+0001xx00 AVESTAN LETTER TH
 U+0001xx00 AVESTAN LETTER TT
 U+0001xx00 AVESTAN LETTER P
 U+0001xx00 AVESTAN LETTER B
 U+0001xx00 AVESTAN LETTER W
 U+0001xx00 AVESTAN LETTER F
 U+0001xx00 AVESTAN LETTER NG
 U+0001xx00 AVESTAN LETTER NNG
 U+0001xx00 AVESTAN LETTER N
 U+0001xx00 AVESTAN LETTER NN
 U+0001xx00 AVESTAN LETTER M
 U+0001xx00 AVESTAN LETTER INITIAL Y
 U+0001xx00 AVESTAN LETTER Y
 U+0001xx00 AVESTAN LETTER INITIAL V
 U+0001xx00 AVESTAN LETTER V
 U+0001xx00 AVESTAN LETTER R
 U+0001xx00 AVESTAN LETTER S
 U+0001xx00 AVESTAN LETTER Z
 U+0001xx00 AVESTAN LETTER SH
 U+0001xx00 AVESTAN LETTER SHH
 U+0001xx00 AVESTAN LETTER SSH
 U+0001xx00 AVESTAN LETTER ZH
 U+0001xx00 AVESTAN LETTER H
 U+0001xx00 AVESTAN LETTER HH
 U+0001xx00 AVESTAN LETTER XV
 U+0001xx00 AVESTAN WORD BREAK
 U+0001xx00 AVESTAN SEMICOLON
 U+0001xx00 AVESTAN COLON
 U+0001xx00 AVESTAN STOP
 U+0001xx00 AVESTAN HYPHEN
 U+0001xx00 (This position shall not be used)
 U+0001xx00 (This position shall not be used)
 U+0001xx00 (This position shall not be used)
 U+0001xx00 AVESTAN NUMBER ONE
 U+0001xx00 AVESTAN NUMBER TWO
 U+0001xx00 AVESTAN NUMBER THREE
 U+0001xx00 AVESTAN NUMBER FOUR
 U+0001xx00 AVESTAN NUMBER TEN
 U+0001xx00 AVESTAN NUMBER TWENTY
 U+0001xx00 AVESTAN NUMBER FORTY

	xx0	xx1	xx2	xx3
0	𐬀	𐬁	𐬂	·
1	𐬃	𐬄	𐬅	·
2	𐬇	𐬈	𐬉	·
3	𐬋	𐬌	𐬍	𐬎
4	𐬏	𐬐	𐬑	—
5	𐬒	𐬓	𐬔	
6	𐬕	𐬖	𐬗	
7	𐬙	𐬚	𐬛	
8	𐬜	𐬝	𐬞	𐬟
9	𐬠	𐬡	𐬢	𐬣
A	𐬤	𐬥	𐬦	𐬧
B	𐬨	𐬩	𐬪	𐬫
C	𐬬	𐬭	𐬮	𐬯
D	𐬰	𐬱	𐬲	𐬳
E	𐬵	𐬶	𐬷	𐬸
F	𐬹	𐬺	𐬻	𐬼

U+0001xx00

AVESTAN NUMBER ONE THOUSAND

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

Unicode Consortium. 1992. *Unicode Technical Report #3: exploratory proposals*

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