







Doc Type: Working Group Document
Title: Proposal to Encode the Samvat Date Sign for Arabic in ISO/IEC 10646
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2009-11-20

1 Introduction


This is a proposal to encode a new character for the Arabic script:  ARABIC SIGN SAMVAT. The character is used in Urdu orthography for writing dates of the Śaka *saṃvat* (سَمَوْتِ *samvat*, from Sanskrit संवत् *saṃvat* < संवत्सर *saṃvatsara* ‘year’), or calendar. It is functionally similar to  U+0601 ARABIC SIGN SANAH (Arabic سنة *sanah* ‘year’). However, ARABIC SIGN SANAH is a generic date marker, while ARABIC SIGN SAMVAT is used specifically for writing dates of the Śaka era.

The Śaka is one of three major calendars used in India. The other two are the Islamic and Gregorian calendars. Legend holds that the Sātavāhana king Śālivāhana established the Śaka era after his victory over the king Vikramāditya in 78 CE, which is the zero year for the era. According to another legend, Vikramāditya is the founder of another popular calendar, the Vikrama *saṃvat*, which was established in 56 BCE. The Śaka, or more accurately the Śālivāhana Śaka *saṃvat*, is the basis for the Indian national calendar, which was established on Caitra 1, Śaka 1879 (March 22, 1957).

2 Glyph Shape

The glyph  for ARABIC SIGN SAMVAT is based upon the form  shown in Figure 1. It is drawn in the Naskh style, rather than in the Nastaliq form of the original, in order to conform to the design of the Arabic font used in the code charts. The glyph represents a stylized abbreviation of the word ‘سَمَوْتِ’ *samvat*, which consists of the initial form of  U+0633 ARABIC LETTER SEEN, the medial form of  U+0645 ARABIC LETTER MEEM, and a horizontal swash. It is analogous to the abbreviation of संवत् *saṃvat* as सं० *saṃ* in Devanagari.


3 Orthography

The  ARABIC SIGN SAMVAT is a subtending character. The date is written inside the character, to the left of the SEEN-MEEM element and above the horizontal stroke. For example, the current Gregorian year 2009 CE corresponds to Śaka 1931 and may be represented in Urdu as سہ ۱۹۳۱.

In Urdu orthography, the era is generally indicated when dates are written. Years of the Islamic calendar are denoted with an abbreviation of the Arabic term ہجری *hijrī* ‘flight of Muhammad from Mecca to Medina’, represented as ه U+06BE ARABIC LETTER HEH DOACHASHMEE, written after the year: ه ۱۴۳۰ ‘1430 AH’ (= 2009 CE). The Gregorian era is marked using an abbreviation of the Arabic عیسوی *īṣawī* ‘Christian’, represented as ع, the initial form of ع U+0639 ARABIC LETTER AIN, also written after the year: ع ۲۰۰۹ ‘2009 CE’. These abbreviations are also used with ARABIC SIGN SANAH, eg. ه ۱۴۳۰ ‘1430 AH’ and ع ۲۰۰۹ ‘2009 CE’. There is no similar convention for indicating years of the Śaka calendar.

4 Implementation

Allocation The character should be encoded at U+0604, in proximity to other Arabic subtending marks.

Representation The ARABIC SIGN SAMVAT should be represented in the code chart as , similar to other subtending marks.

Rendering The swash portion of the glyph is the only subtending element. The shape of the SEEN-MEEM element is static. The length of the subtending swash is dependent upon calligraphic preferences.

Encoding Order In handwriting, the character is generally written after the digits that it is intended to span. This is on account of typographical justifications, such that a writer can produce a swash that spans the digits appropriately. The encoded order is the reverse, that is, the character precedes the sequence of digits that it is intended to span.

Properties The character has the following general properties, expressed in the format used in the Unicode Character Database (`UnicodeData.txt`):

```
0604;ARABIC SIGN SAMVAT;Cf;0;AN;;;;;N;;;;;
```

It has the following shaping properties, expressed in the format used in the Arabic shaping and joining properties database (`ArabicShaping.txt`):

```
0604; ARABIC SIGN SAMVAT; U; No_Joining_Group
```

5 References

Ahmad Dihlavī, Sayyid. 1974. *فرہنگ آصفیہ* [*Farhang-i Āsafiyah*]. vol. 2. Delhi: Taraqqī-e-Urdū Board, National Academy.

6 Acknowledgments

The author would like to thank Roozbeh Pournader for reviewing an earlier version of this proposal and for providing information on character shaping properties and encoding order.

The glyph for ARABIC SIGN SAMVAT was designed with the assistance of Nadia S. Hasan.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

۶۱۹۷۴

۱۸۹۶ء

FARHANG-E-ASIFIA-Vol.II

ترقی اردو بورڈ ایڈیشن

تعداد — تین ہزار

تقسیم کار: نیشنل اکاڈمی

۹- انصاری مارکیٹ، دریا گنج، دہلی ۱۱۰۰۰۶

قیمت جلد دوم: پچیس روپے

نیشنل اکاڈمی دہلی نے ترقی اردو بورڈ، وزارت تعلیم و سماجی بہبود، حکومت ہند کے لیے جید پریس، بیہاران، دہلی سے چھپوا کر شایع کیا۔

Figure 1: Title page of *Farhang-i Āsafīyyah* showing use of ARABIC SIGN SAMVAT (from Ahmad Dihlavī 1974). The date in the boxed text is ۱۸۹۶ء ‘1896 samvat’, which corresponds to ۱۹۷۴ء ‘1974’ of the Gregorian (‘īṣawī) calendar, as shown above the boxed text.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P & P) from
<http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.
Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.
See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title: **Proposal to Encode the Samvat Date Sign for Arabic in ISO/IEC 10646**
2. Requester's name: **University of California, Berkeley Script Encoding Initiative (Universal Scripts Project); author: Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Liaison contribution**
4. Submission date: **2009-11-20**
5. Requester's reference (if applicable): **N/A**
6. Choose one of the following:
 - (a) This is a complete proposal: **Yes**
 - (b) or, More information will be provided later: **No**

B. Technical - General

1. Choose one of the following:
 - (a) This proposal is for a new script (set of characters): **No**
 - i. Proposed name of script: **N/A**
 - (b) The proposal is for addition of character(s) to an existing block: **Yes**
 - i. Name of the existing block: **Arabic**
2. Number of characters in proposal: **1**
3. Proposed category: **B.1 - Specialized (small collection)**
4. Is a repertoire including character names provided?: **Yes**
 - (a) If Yes, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?: **Yes**
 - (b) Are the character shapes attached in a legible form suitable for review?: **Yes**
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?: **Anshuman Pandey; True Type**
 - (a) If available now, identify source(s) for the font and indicate the tools used: **The glyph was designed in FontForge by Anshuman Pandey.**
6. References:
 - (a) Are references (to other character sets, dictionaries, descriptive texts etc.) provided?: **Yes**
 - (b) Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?: **Yes**
7. Special encoding issues:
 - (a) 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; see text of the proposal.**
8. 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. 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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard. **Character properties are included.**

¹ Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03)

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 (for example: National Body, user groups of the script or characters, other experts, etc.)? **No**
 - (a) If Yes, with whom?: **N/A**
 - i. If Yes, available relevant documents: **N/A**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**
 - (a) Reference: **The character is used in specialized Urdu typography in India.**
4. The context of use for the proposed characters (type of use; common or rare): **Rare**
 - (a) Reference: **The character is a specialized sign used for writing dates in Urdu.**
5. Are the proposed characters in current use by the user community?: **Yes.**
 - (a) If Yes, where? Reference: **India.**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?: **No**
 - (a) If Yes, is a rationale provided?: **N/A**
 - i. If Yes, reference: **N/A**
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **No**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**
 - (a) If Yes, is a rationale for its inclusion provided?: **N/A**
 - i. If Yes, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
 - (a) If Yes, is a rationale provided?: **N/A**
 - i. If Yes, reference: **N/A**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? **No**
 - (a) If Yes, is a rationale for its inclusion provided? **N/A**
 - i. If Yes, reference: **N/A**
11. 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**
 - (a) If Yes, is a rationale for such use provided? **N/A**
 - i. If Yes, reference: **N/A**
 - (b) Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **No**
 - i. If Yes, reference: **N/A**
12. Does the proposal contain characters with any special properties such as control function or similar semantics? **No**
 - (a) If Yes, describe in detail (include attachment if necessary): **N/A**
13. Does the proposal contain any Ideographic compatibility character(s)? **No**
 - (a) If Yes, is the equivalent corresponding unified ideographic character(s) identified? **N/A**
 - i. If Yes, reference: **N/A**