Title: Final Proposal to Encode Coptic Numbers in the UCS
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2010-05-24

1 Introduction

This is a proposal to encode in the Universal Character Set (UCS) a set of characters used for writing numbers in Coptic. It builds upon and replaces the following documents:

- N3786 L2/10-114 "Towards an Encoding for Coptic Numbers in the UCS"
- L2/09-163R "Proposal to Encode Coptic Numerals in ISO/IEC 10646"

Several changes to the original document have been made, including change of name of the script block from "Coptic Numerals" to "Coptic Numbers" and allocation of the block in the Supplementary Multilingual Plane (SMP). A font has also been developed.

2 Background

The proposed characters are elements of a numeric notation system used in some Coptic manuscripts, which differ from the standard representation of numbers using letters of the alphabet. A comparison of the two notation systems is given in Table 1. These Coptic numbers are regarded as 'cursive' forms of ordinary Coptic letters. In an illustration in *Grammaire Copte* (1956), Alexis Mallon suggests the transformation of Coptic letters into distinct numbers, which he calls 'cursive Coptic numbers' ('chiffres coptes cursifs'). It is likely that these numbers were used primarily in Coptic-Arabic manuscripts, such as that shown in Figure 2. They appear in Coptic manuscript fragments, such as those held in the collection of the AHRC Rylands Cairo Genizah Project at the University of Manchester (Figure 3).

The Coptic Numbers appear in specimens included by Michael Everson in "Revised proposal to add the Coptic alphabet to the BMP of the UCS" (N2636), ie. in Figure 14, of which an excerpt is given here in Figure 6. Everson, however, did not propose the encoding of these characters, but stated that "further study may indicate that some of the additional characters and symbols shown here should also be added to the Standard". Indeed, additional research has shown that the 'Signes de numération' illustrated in the figure were also described by Antoine P. Pihan in *Exposé des signes de numération* (1860). This proposal seeks to continue Everson's work of developing support for Coptic in the UCS.

As the Coptic Numbers may be variations on the standard cursive forms of Coptic letters, it may be possible to unify them with existing Coptic letters. However, the depiction of these characters as unique elements of the Coptic script, as shown in Figure 4 and Figure 6, suggests that they were considered sufficiently distinct from the original alphabetic sources. These factors strongly suggest an independent encoding for Coptic Numbers. An encoding for the Coptic Numbers will enhance the Coptic repertoire by offering a means for representing characters used in Coptic-Arabic manuscripts.

3 Characters Proposed

The characters are proposed for encoding in a new script block to be named 'Coptic Numbers'. The 29 characters are allocated in the SMP at the range U+102E0..U+102FF. The proposed code chart and names list are shown in Figure 1.

The names of the characters follow UCS naming conventions. Digits 1–9 are named DIGIT and all other numbers are named NUMBER.

4 The Notation System

Structure Coptic Numbers represent units of a positional decimal system. It is an additive system, in which the value of a numeric sequence is the sum of the values of the numbers that constitute it. Number are written left-to-right. There is no character for zero; it is inherently represented in the distinct number for each decimal order. The system has unique characters for representing decimal orders of the primary, tens, and hundreds units. The thousands are represented by writing a slash-mark beneath the numbers.

Orthography The thousands are represented by writing the primary number and ◌ THOUSANDS MARK: ⲉ FIVE + ◌ THOUSANDS MARK = ⲉ 5,000. The ten thousands are written using the tens number and the THOUSANDS MARK: ⲛ FIFTY + ◌ THOUSANDS MARK = ⲛ 50,000. The hundred thousands are written with the numbers for the hundreds and the THOUSANDS MARK: ⳟ FIVE HUNDRED + ◌ THOUSANDS MARK = ⳟ 500,000.

In theory, decimal orders larger than hundred thousand may be represented by writing the THOUSANDS MARK twice, eg. ⲇ = 1,000; ⲇ = 1,000,000. This practice mirrors the principle of writing numbers using letters of the Coptic alphabet, in which the overline ◌̅ U+0305 COMBINING OVERLINE is doubled (◌̿ U+033F COMBINING DOUBLE OVERLINE) to indicate the order of the thousands, eg. ⲁ̅ = 1; ⲁ̿ = 1,000.

Composite numbers are produced using the primary numbers and the numbers of larger decimal orders. The larger numeral is written first, then the primary numeral: ⲱⲉ = 25 (TWENTY + FIVE); ⲥⲉ = 205 (TWO HUNDRED + FIVE); ⲥⲛ = 250 (TWO HUNDRED + FIFTY).

Numbers are marked using the ◌͠ COPTIC NUMBER MARK. The length of the NUMBER MARK extends over the entire sequence of numbers: ⲓ͠ⲉ = 15; ⳟⲛ͠ = 550; ⲉⳟⲉ͠ = 5,505.

5 Character Properties

The characters of the Coptic Numbers block have the following properties:

```
102E0 COPTIC THOUSANDS MARK;Mn;0;NSM;;;;1000;N;;;;;
102E1 COPTIC DIGIT ONE;No;0;L;;;;1;N;;;;;
102E2 COPTIC DIGIT TWO;No;0;L;;;;2;N;;;;;
102E3 COPTIC DIGIT THREE;No;0;L;;;;3;N;;;;;
102E4 COPTIC DIGIT FOUR;No;0;L;;;;4;N;;;;;
102E5 COPTIC DIGIT FIVE;No;0;L;;;;5;N;;;;;
102E6 COPTIC DIGIT SIX;No;0;L;;;;6;N;;;;;;
102E7 COPTIC DIGIT SEVEN;No;0;L;;;;7;N;;;;;
102E8 COPTIC DIGIT EIGHT;No;0;L;;;;8;N;;;;;
102E9 COPTIC DIGIT NINE;No;0;L;;;;9;N;;;;;
102EA COPTIC NUMBER TEN;No;0;L;;;;10;N;;;;;
102EB COPTIC NUMBER TWENTY;No;0;L;;;;20;N;;;;;
102EC COPTIC NUMBER THIRTY;No;0;L;;;;30;N;;;;;
```

```
102ED COPTIC NUMBER FORTY;No;0;L;;;;40;N;;;;;
102EE COPTIC NUMBER FIFTY;No;0;L;;;;50;N;;;;;
102EF COPTIC NUMBER SIXTY;No;0;L;;;;60;N;;;;;
102F0 COPTIC NUMBER SEVENTY;No;0;L;;;;70;N;;;;;
102F1 COPTIC NUMBER EIGHTY;No;0;L;;;;80;N;;;;;
102F2 COPTIC NUMBER NINETY;No;0;L;;;;90;N;;;;;
102F3 COPTIC NUMBER ONE HUNDRED;No;0;L;;;;100;N;;;;;
102F4 COPTIC NUMBER TWO HUNDRED;No;0;L;;;;200;N;;;;;
102F5 COPTIC NUMBER THREE HUNDRED;No;0;L;;;;300;N;;;;;
102F6 COPTIC NUMBER FOUR HUNDRED;No;0;L;;;;400;N;;;;;
102F7 COPTIC NUMBER FIVE HUNDRED;No;0;L;;;;500;N;;;;;
102F8 COPTIC NUMBER SIX HUNDRED;No;0;L;;;;600;N;;;;;
102F9 COPTIC NUMBER SEVEN HUNDRED;No;0;L;;;;700;N;;;;;;
102FA COPTIC NUMBER EIGHT HUNDRED;No;0;L;;;;800;N;;;;;
102FB COPTIC NUMBER NINE HUNDRED;No;0;L;;;;900;N;;;;;
102FC COPTIC NUMBER MARK;Cf;0;L;;;;;N;;;;;
```

6 References

AHRC Rylands Cairo Genizah Project. Rylands Genizah fragment B 6548-1. http://enriqueta.man.ac.uk:8180/luna/servlet/ManchesterDev~95~2. Accessed April 2010.

Everson, Michael. 2003. "Revised proposal to add the Coptic alphabet to the BMP of the UCS" (ISO/IEC JTC1/SC2/WG2 N2636). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2636.pdf

King, David A. 2001. *The Ciphers of the Monks: A Forgotten Number-Notation of the Middle Ages*. Stuttgart: F. Steiner.

Mallon, Alexis. 1956. *Grammaire Copte*. Bibliographie, chrestomathie et vocabulaire. Beyrouth: Imprimerie catholique.

Pandey, Anshuman. 2009. "Proposal to Encode Coptic Numerals in ISO/IEC 10646" (L2/09-163R). http://www-personal.umich.edu/~pandey/09163-copticnumerals.pdf

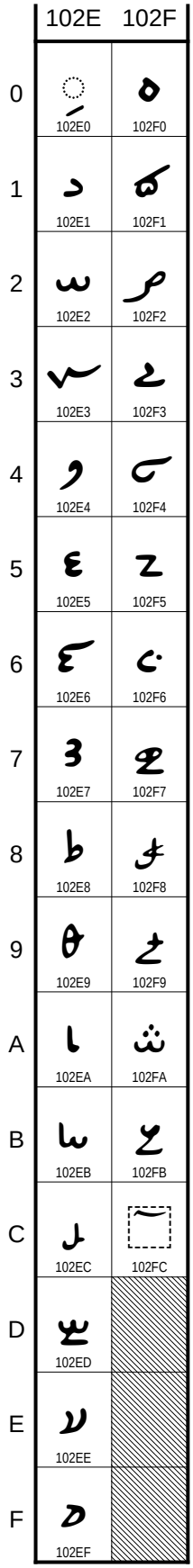
———. 2010. "Towards an Encoding for Coptic Numbers in the UCS" (ISO/IEC JTC1/SC2/WG2 N3786 L2/10-114). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3786.pdf

Pihan, Antoine Paulin. 1860. *Exposé des signes de numération usités chez les peuples orientaux anciens et modernes*. Paris: L'imprimerie impériale.

7 Acknowledgments

I would like to thank Traianos Gagos and Terry Wilfong, both of the University of Michigan, for sharing their comments on L2/09-163R.

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.



Sign

102E0 ◌ COPTIC THOUSANDS MARK

Digits

102E1 COPTIC DIGIT ONE
102E2 COPTIC DIGIT TWO
102E3 COPTIC DIGIT THREE
102E4 COPTIC DIGIT FOUR
102E5 COPTIC DIGIT FIVE
102E6 COPTIC DIGIT SIX
102E7 COPTIC DIGIT SEVEN
102E8 COPTIC DIGIT EIGHT
102E9 COPTIC DIGIT NINE

Numbers

102EA COPTIC NUMBER TEN
102EB COPTIC NUMBER TWENTY
102EC COPTIC NUMBER THIRTY
102ED COPTIC NUMBER FORTY
102EE COPTIC NUMBER FIFTY
102EF COPTIC NUMBER SIXTY
102F0 COPTIC NUMBER SEVENTY
102F1 COPTIC NUMBER EIGHTY
102F2 COPTIC NUMBER NINETY
102F3 COPTIC NUMBER ONE HUNDRED
102F4 COPTIC NUMBER TWO HUNDRED
102F5 COPTIC NUMBER THREE HUNDRED
102F6 COPTIC NUMBER FOUR HUNDRED
102F7 COPTIC NUMBER FIVE HUNDRED
102F8 COPTIC NUMBER SIX HUNDRED
102F9 COPTIC NUMBER SEVEN HUNDRED
102FA COPTIC NUMBER EIGHT HUNDRED
102FB COPTIC NUMBER NINE HUNDRED

Number Mark

102FC ◌ COPTIC NUMBER MARK

Figure 1: Proposed code chart and nameslist for Coptic Numbers

	1	2	3	4	5	6	7	8	9
1	𐋡	𐋢	𐋣	𐋤	𐋥	𐋦	𐋧	𐋨	𐋩
	ⲁ̅	ⲃ̅	ⲅ̅	ⲇ̅	ⲉ̅	ⲋ̅	ⲍ̅	ⲏ̅	ⲑ̅
10	𐋪	𐋫	𐋬	𐋭	𐋮	𐋯	𐋰	𐋱	𐋲
	ⲓ̅	ⲕ̅	ⲗ̅	ⲙ̅	ⲛ̅	ⲝ̅	ⲟ̅	ⲡ̅	ϥ̅
100	𐋳	𐋴	𐋵	𐋶	𐋷	𐋸	𐋹	𐋺	𐋻
	ⲣ̅	ⲥ̅	ⲧ̅	ⲩ̅	ⲫ̅	ⲭ̅	ⲯ̅	ⲱ̅	ⳁ̅
1,000	𐋡𐋠	𐋢𐋠	𐋣𐋠	𐋤𐋠	𐋥𐋠	𐋦𐋠	𐋧𐋠	𐋨𐋠	𐋩𐋠
	ⲁ̿	ⲃ̿	ⲅ̿	ⲇ̿	ⲉ̿	ⲋ̿	ⲍ̿	ⲏ̿	ⲑ̿
10,000	𐋪𐋠	𐋫𐋠	𐋬𐋠	𐋭𐋠	𐋮𐋠	𐋯𐋠	𐋰𐋠	𐋱𐋠	𐋲𐋠
	ⲓ̿	ⲕ̿	ⲗ̿	ⲙ̿	ⲛ̿	ⲝ̿	ⲟ̿	ⲡ̿	ϥ̿
100,000	𐋳𐋠	𐋴𐋠	𐋵𐋠	𐋶𐋠	𐋷𐋠	𐋸𐋠	𐋹𐋠	𐋺𐋠	𐋻𐋠
	ⲣ̿	ⲥ̿	ⲧ̿	ⲩ̿	ⲫ̿	ⲭ̿	ⲯ̿	ⲱ̿	ⳁ̿

Table 1: Numbers written using Coptic Numbers (top) and the alphabetic system (bottom).

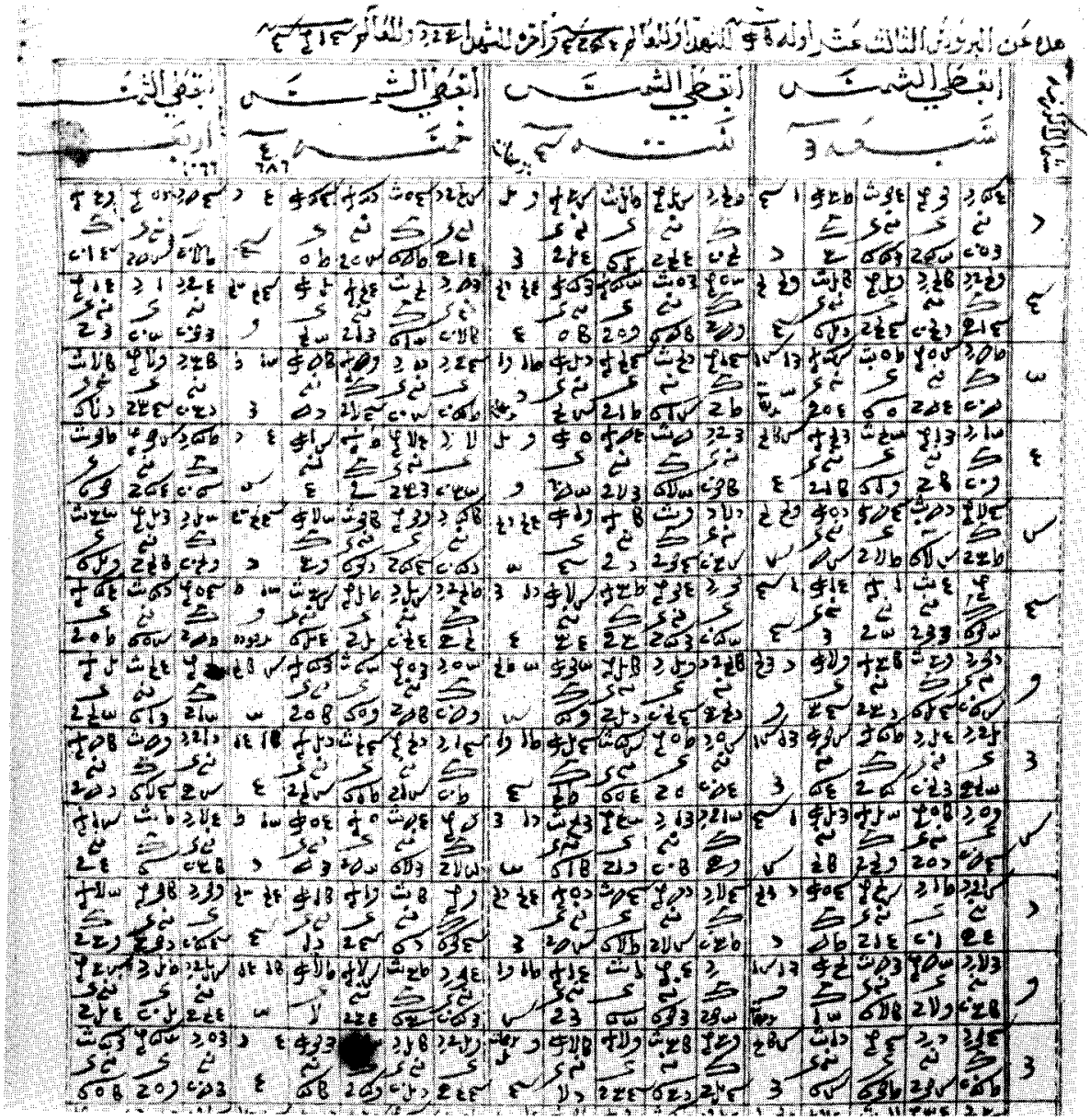


Fig. C.6 Coptic numerals in a copy from *ca.* 1800 of a set of astronomical tables by the early-13th-century Coptic scholar Ibn ‘Assal. This notation has separate, unrelated symbols for the units, tens and hundreds, *etc.*, and for simple fractions. (From MS Cairo DM 910,1, fol. 81v, courtesy of the Egyptian National Library.)

Figure 2: Excerpt of an astronomical table showing the use of Coptic Numbers with the Arabic script (from King 2001: Appendix C, p. 299).

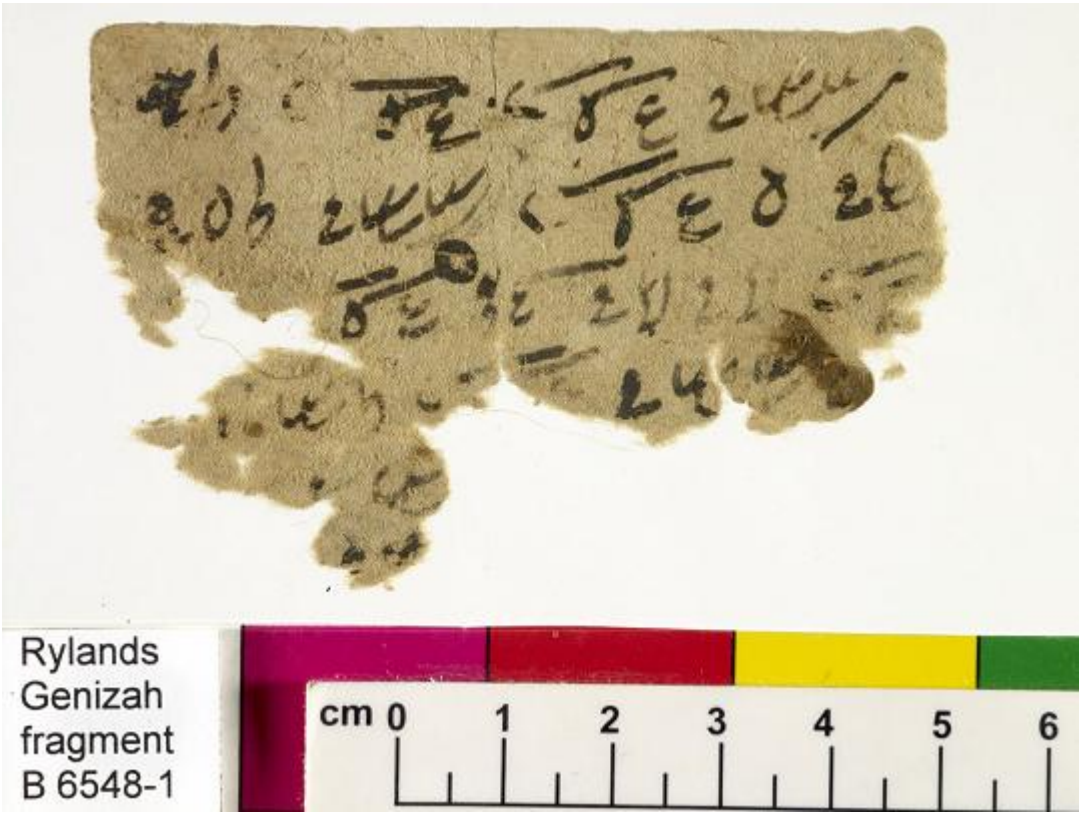


Figure 3: Coptic numbers in a manuscript fragment from the Rylands Genizah collection (from AHRC Rylands Cairo Genizah Project: fragment B 6548-1).

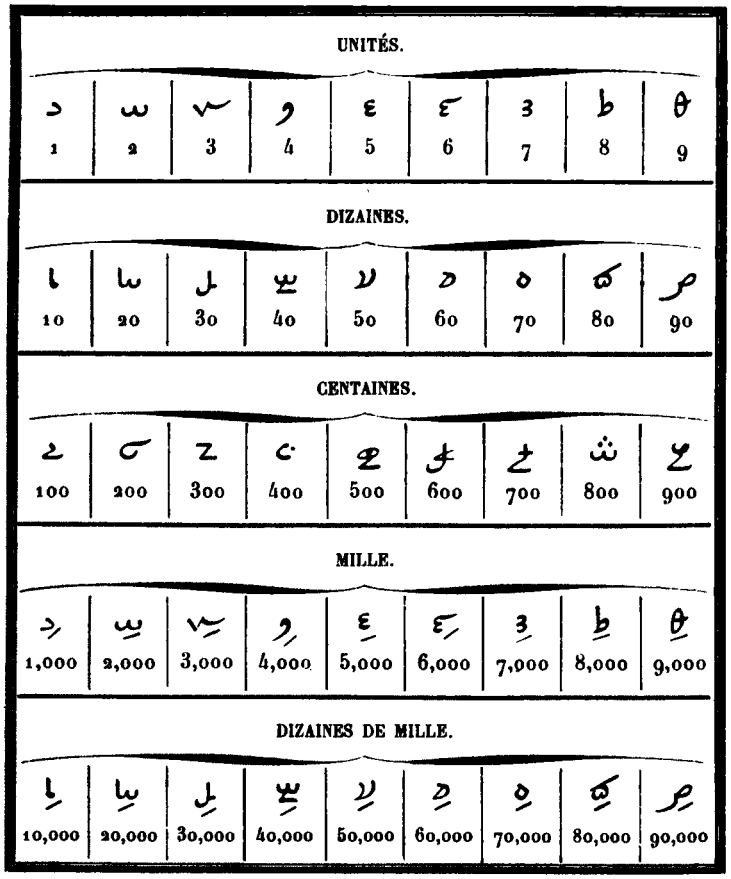


Figure 4: Table showing the forms of Coptic Numbers (from Pihan 1860: 213). Compare the Coptic Numbers to the alphabetic system system shown in Figure 7.

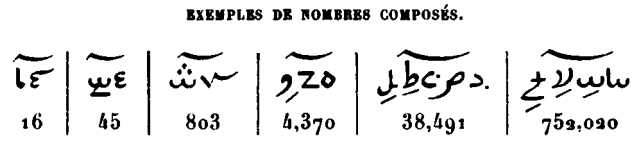


Figure 5: Table showing composite numbers written with Coptic Numbers (from Pihan 1860: 214).

Figure 6: Coptic Numbers for the primary, tens, hundreds, and thousands shown in a specimen of Coptic type under the heading 'Signes de numération'. Two length variants of the COPTIC NUMBER MARK are shown under the heading 'Lettres accentuées...' (reproduced from Everson 2003: Figure 14).

ⲁ̅ = 1	ⲕ̅ = 20	ⲧ̅ = 300
ⲃ̅ = 2	ⲗ̅ = 30	ⲩ̅ = 400
ⲅ̅ = 3	ⲙ̅ = 40	ⲫ̅ = 500
ⲇ̅ = 4	ⲛ̅ = 50	ⲭ̅ = 600
ⲉ̅ = 5	ⲝ̅ = 60	ⲯ̅ = 700
ⲋ̅ = 6	ⲟ̅ = 70	ⲱ̅ = 800
ⲍ̅ = 7	ⲡ̅ = 80	ⳡ̅ = 900
ⲏ̅ = 8	ⲣ̅ = 90	ⲁ̿ = 1.000
ⲑ̅ = 9	ⲥ̅ = 100	ⲃ̿ = 2.000
ⲓ̅ = 10	ⲧ̅ = 200	ⲓ̿ = 10.000

Figure 7: The representation of numbers in Coptic using letters of the alphabet and horizontal overlines (reproduced from Everson 2003: Figure 12).

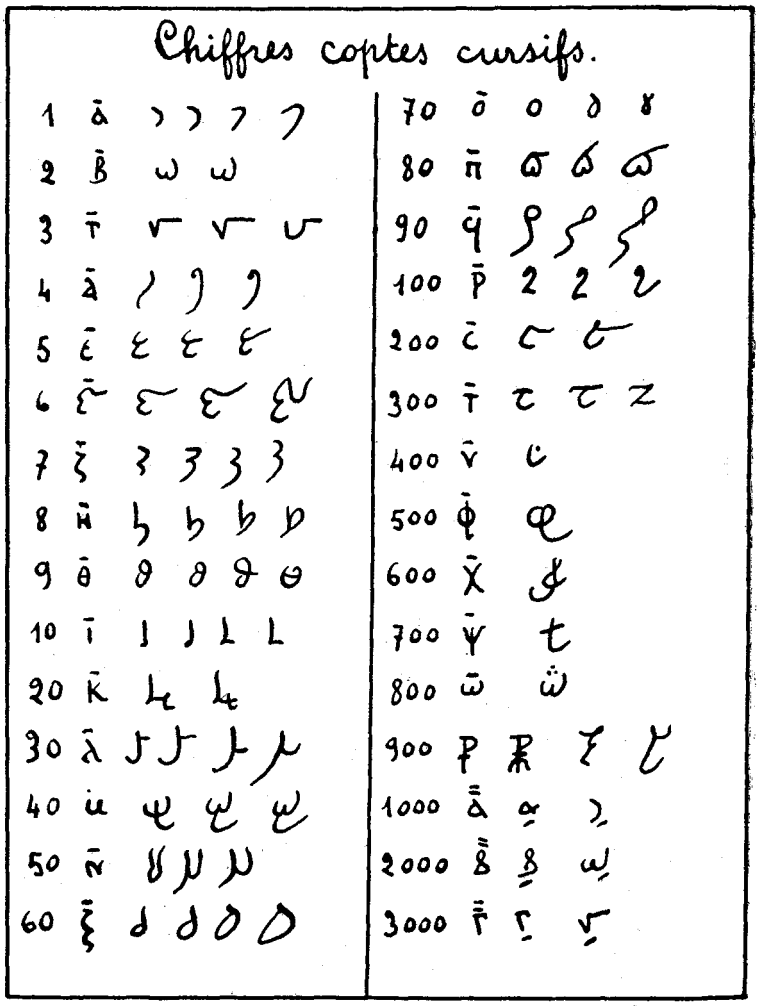


Figure 8: Table showing the Coptic numbers (from Mallon 1956: 234).

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]

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: **Final Proposal to Encode Coptic Numbers in the UCS**
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: **2010-05-24**
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): **Yes**
      i. Proposed name of script: **Coptic Numbers**
   (b) The proposal is for addition of character(s) to an existing block: **No**
      i. Name of the existing block: **N/A**
2. Number of characters in proposal: **29**
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 font was designed by Anshuman Pandey using FontForge.**
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.)? **Yes**
   (a) If Yes, with whom?: **Traianos Gagos and Terry Wilfong (University of Michigan)**
      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; see text of the proposal.**
   (a) Reference: **N/A**
4. The context of use for the proposed characters (type of use; common or rare): **Common**
   (a) Reference: **The characters were used in Coptic Arabic manuscripts.**
5. Are the proposed characters in current use by the user community?: **No.**
   (a) If Yes, where? Reference: **N/A**
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)? **Yes; the characters belong to a set.**
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**