

Title: Proposal to Encode Coptic Numerals in ISO/IEC 10646
Source: Anshuman Pandey (pandey@umich.edu) / Script Encoding Initiative (SEI)
Status: Liaison Contribution
Action: For consideration by UTC
Date: 2009-09-15

1 Introduction

This is a proposal to encode Coptic Numerals in the Universal Character Set (ISO/IEC 10646). The characters proposed here are elements of a Coptic numeric notation system that differs from the representation of numbers using letters from the Coptic alphabet (see Table 1 for a comparison of the two systems). The Coptic Numerals belong to a notation system has unique numerals for the primary, tens, and hundreds units. The thousands are represented by writing a slash-mark beneath the numerals.

The Coptic Numerals were not included by Michael Everson in his “Revised proposal to add the Coptic alphabet to the BMP of the UCS” (N2636). The Coptic Numerals proposed here appear in Figure 14 of N2636, an excerpt of which is given in Figure 4. In the caption of the figure, Everson writes 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 her *Exposé des signes de numération* (see Figure 2). Encoding these characters in the UCS will enable representation of the historical numeric notation system of Coptic.

2 Characters Proposed

There is a requirement of 29 characters for fully representing the Coptic Numerals:

Primary Numerals Nine characters:

ⲁ	COPTIC NUMERAL ONE	ϥ	COPTIC NUMERAL SIX
Ⲃ	COPTIC NUMERAL TWO	ⲛ	COPTIC NUMERAL SEVEN
Ⲅ	COPTIC NUMERAL THREE	Ⲕ	COPTIC NUMERAL EIGHT
Ⲇ	COPTIC NUMERAL FOUR	ⲕ	COPTIC NUMERAL NINE
Ⲉ	COPTIC NUMERAL FIVE		

Tens Numerals Nine characters:

Ⲑ	COPTIC NUMERAL TEN	Ⲛ	COPTIC NUMERAL SIXTY
ⲑ	COPTIC NUMERAL TWENTY	ⲛ	COPTIC NUMERAL SEVENTY
Ⲓ	COPTIC NUMERAL THIRTY	Ⲕ	COPTIC NUMERAL EIGHTY
ⲓ	COPTIC NUMERAL FORTY	ⲕ	COPTIC NUMERAL NINETY
Ⲕ	COPTIC NUMERAL FIFTY		

Hundreds Numerals Nine characters:

Ⲛ	COPTIC NUMERAL ONE HUNDRED	Ⲛ	COPTIC NUMERAL SIX HUNDRED
ⲛ	COPTIC NUMERAL TWO HUNDRED	ⲛ	COPTIC NUMERAL SEVEN HUNDRED
Ⲝ	COPTIC NUMERAL THREE HUNDRED	Ⲝ	COPTIC NUMERAL EIGHT HUNDRED
ⲝ	COPTIC NUMERAL FOUR HUNDRED	ⲝ	COPTIC NUMERAL NINE HUNDRED
Ⲟ	COPTIC NUMERAL FIVE HUNDRED		

Various Signs Two characters:

Ⲑ	COPTIC THOUSANDS MARK	ⲑ	COPTIC NUMBER MARK
---	-----------------------	---	--------------------

3 The Notation System

Structure Coptic Numerals represent units of a base-10 (decimal) positional system. The notation system is additive, that is, the value of a number is the sum of the values of the numerals that constitute it. There is no character for zero; it is inherently represented in the distinct numerals for the various decimal orders. The numerals are written left-to-right.

Orthography The thousands are represented by writing the primary numerals and Ⲑ COPTIC THOUSANDS MARK: ⲉ FIVE + Ⲑ THOUSANDS MARK = ⲉⲐ 5,000. The ten thousands are written using the tens numerals and the THOUSANDS MARK: ⲙ FIFTY + Ⲑ THOUSANDS MARK = ⲙⲐ 50,000. The hundred thousands are written with the numerals for the hundreds and the THOUSANDS MARK: Ⲛ FIVE HUNDRED + Ⲑ THOUSANDS MARK = ⲚⲐ 500,000.

Numbers of decimal orders larger than the hundred thousands may theoretically be represented by multiple use of the THOUSANDS MARK, however, this practice is unattested. This practice is similar in principle to the writing of numbers using letters of the Coptic alphabet, where the overline is doubled to indicate the orders of the thousands.

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

Numbers are marked using the ⲑ COPTIC NUMBER MARK. The length of the NUMBER MARK extends to cover the number of numerals written. ⲑⲉ = 15; ⲑⲙ = 550; ⲑⲚⲉ = 5,505.

4 Implementation**4.1 Allocation**

There is no allocation for the Coptic Numerals in the Unicode Roadmaps. Also, there is no space in the Coptic block (U+2C80..2CFF) or in the the Greek and Coptic block (U+0370..03FF). An allocation for Coptic Numerals requires a minimum of 29 code-points. Some space should be reserved to account for fraction signs and other characters that may be identified through further investigation. The range U+102E0..102FF in the Supplementary Multilingual Plane (SMP) is a possible location for Coptic Numerals, offering 32 code-points.

4.2 Character Properties

The Coptic Numerals have the following properties, expressed in the format of the Unicode Character Database:

```

xx01 COPTIC NUMERAL ONE;No;0;L;;;1;N;;;;;
xx02 COPTIC NUMERAL TWO;No;0;L;;;2;N;;;;;
xx03 COPTIC NUMERAL THREE;No;0;L;;;3;N;;;;;
xx04 COPTIC NUMERAL FOUR;No;0;L;;;4;N;;;;;
xx05 COPTIC NUMERAL FIVE;No;0;L;;;5;N;;;;;
xx06 COPTIC NUMERAL SIX;No;0;L;;;6;N;;;;;
xx07 COPTIC NUMERAL SEVEN;No;0;L;;;7;N;;;;;
xx08 COPTIC NUMERAL EIGHT;No;0;L;;;8;N;;;;;
xx09 COPTIC NUMERAL NINE;No;0;L;;;9;N;;;;;
xx0A COPTIC NUMERAL TEN;No;0;L;;;10;N;;;;;
xx0B COPTIC NUMERAL TWENTY;No;0;L;;;20;N;;;;;
xx0C COPTIC NUMERAL THIRTY;No;0;L;;;30;N;;;;;
xx0D COPTIC NUMERAL FORTY;No;0;L;;;40;N;;;;;
xx0E COPTIC NUMERAL FIFTY;No;0;L;;;50;N;;;;;
xx0F COPTIC NUMERAL SIXTY;No;0;L;;;60;N;;;;;
xx10 COPTIC NUMERAL SEVENTY;No;0;L;;;70;N;;;;;
xx11 COPTIC NUMERAL EIGHTY;No;0;L;;;80;N;;;;;
xx12 COPTIC NUMERAL NINETY;No;0;L;;;90;N;;;;;
xx13 COPTIC NUMERAL ONE HUNDRED;No;0;L;;;100;N;;;;;
xx14 COPTIC NUMERAL TWO HUNDRED;No;0;L;;;200;N;;;;;
xx15 COPTIC NUMERAL THREE HUNDRED;No;0;L;;;300;N;;;;;
xx16 COPTIC NUMERAL FOUR HUNDRED;No;0;L;;;400;N;;;;;
xx17 COPTIC NUMERAL FIVE HUNDRED;No;0;L;;;500;N;;;;;
XX18 COPTIC NUMERAL SIX HUNDRED;No;0;L;;;600;N;;;;;
XX19 COPTIC NUMERAL SEVEN HUNDRED;No;0;L;;;700;N;;;;;
xx2A COPTIC NUMERAL EIGHT HUNDRED;No;0;L;;;800;N;;;;;
xx2B COPTIC NUMERAL NINE HUNDRED;No;0;L;;;900;N;;;;;
XX2C COPTIC THOUSANDS MARK;Mn;0;NSM;;;1000;N;;;;;
XX2D COPTIC NUMBER MARK;So;0;ET;;;N;;;;;

```

5 References

- Everson, Michael. 2003. “Revised proposal to add the Coptic alphabet to the BMP of the UCS.” ISO/IEC JTC1/SC2/WG2 N2636. October 1, 2003. <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.
- Pihan, Antoine Paulin. 1860. *Exposé des signes de numération usités chez les peuples orientaux anciens et modernes*. Paris: L’imprimerie impériale.

6 Acknowledgments

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.

	1	2	3	4	5	6	7	8	9
1	ⲁ ⲁ̅	Ⲃ Ⲃ̅	ⲃ ⲃ̅	Ⲅ Ⲅ̅	ⲅ ⲅ̅	Ⲇ Ⲇ̅	ⲇ ⲇ̅	Ⲉ Ⲉ̅	ⲉ ⲉ̅
10	Ⲋ Ⲋ̅	ⲋ ⲋ̅	Ⲍ Ⲍ̅	ⲍ ⲍ̅	Ⲏ Ⲏ̅	ⲏ ⲏ̅	Ⲑ Ⲑ̅	ⲑ ⲑ̅	Ⲓ Ⲓ̅
100	Ⲕ Ⲕ̅	ⲕ ⲕ̅	Ⲗ Ⲗ̅	ⲗ ⲗ̅	Ⲙ Ⲙ̅	ⲙ ⲙ̅	Ⲛ Ⲛ̅	ⲛ ⲛ̅	Ⲝ Ⲝ̅
1,000	Ⲟ Ⲟ̅	ⲟ ⲟ̅	Ⲡ Ⲡ̅	ⲡ ⲡ̅	Ⲣ Ⲣ̅	ⲣ ⲣ̅	Ⲥ Ⲥ̅	ⲥ ⲥ̅	Ⲧ Ⲧ̅
10,000	ⲧ̅ ⲧ̅̅	Ⲩ̅ Ⲩ̅̅	ⲩ̅ ⲩ̅̅	Ⲫ̅ Ⲫ̅̅	ⲫ̅ ⲫ̅̅	Ⲭ̅ Ⲭ̅̅	ⲭ̅ ⲭ̅̅	Ⲯ̅ Ⲯ̅̅	ⲯ̅ ⲯ̅̅
100,000	Ⲱ̅ Ⲱ̅̅	ⲱ̅ ⲱ̅̅	Ⲳ̅ Ⲳ̅̅	ⲳ̅ ⲳ̅̅	Ⲵ̅ Ⲵ̅̅	ⲵ̅ ⲵ̅̅	Ⲷ̅ Ⲷ̅̅	ⲷ̅ ⲷ̅̅	Ⲹ̅ Ⲹ̅̅

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

[illegible]

Figure 1: Excerpt of a manuscript showing the use of Coptic Numerals with the Arabic script (from King 2001: Appendix C, p. 299).

UNITÉS.								
ⲁ	Ⲃ	ⲃ	Ⲅ	ⲅ	Ⲇ	ⲇ	Ⲉ	ⲉ
1	2	3	4	5	6	7	8	9
DIZAINES.								
Ⲑ	ⲑ	Ⲓ	ⲓ	Ⲕ	ⲕ	Ⲍ	ⲍ	Ⲏ
10	20	30	40	50	60	70	80	90
CENTAINES.								
Ⲛ	ⲛ	Ⲝ	ⲝ	Ⲟ	ⲟ	Ⲡ	ⲡ	Ⲣ
100	200	300	400	500	600	700	800	900
MILLE.								
Ⲅ	ⲅ	Ⲇ	ⲇ	Ⲉ	ⲉ	Ⲋ	ⲋ	Ⲍ
1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000	9,000
DIZAINES DE MILLE.								
Ⲑ	ⲑ	Ⲓ	ⲓ	Ⲕ	ⲕ	Ⲍ	ⲍ	Ⲏ
10,000	20,000	30,000	40,000	50,000	60,000	70,000	80,000	90,000

Figure 2: Table showing the forms of Coptic Numerals (from Pihan 1860: 213). Compare the Coptic Numerals to the alphabetic system shown in Figure 5.

EXEMPLES DE NOMBRES COMPOSÉS.					
ⲥⲁ	ⲥⲛ	ⲥⲛⲥ	ⲥⲛⲥⲛ	ⲥⲛⲥⲛⲥⲛ	ⲥⲛⲥⲛⲥⲛⲥⲛ
16	45	803	4,370	38,491	752,020

Figure 3: Table showing composite numbers written with Coptic Numerals (from Pihan 1860: 214).

COPTE MEMPHITIQUE.	
LETTRES ALPHABÉTIQUES.	
MAJUSCULES.	MINUSCULES.
Ⲁ Ⲃ Ⲅ Ⲇ Ⲉ Ⲋ Ⲍ Ⲏ Ⲑ	ⲁ ⲃ ⲅ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ
Ⲓ Ⲕ Ⲗ Ⲙ Ⲛ Ⲝ Ⲟ Ⲡ Ⲣ	ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ
ⲣ ⲥ Ⲧ Ⲩ Ⲫ Ⲭ Ⲯ Ⲱ Ⲳ	ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ
Ⲵ Ⲷ Ⲹ Ⲻ Ⲽ Ⲿ ⲿ	ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ
ⲱ ⲳ ⲵ ⲷ ⲹ ⲻ ⲽ	ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ
ⲿ ⲱ ⲳ ⲵ ⲷ ⲹ ⲻ ⲽ	ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ
SIGNES DE NUMÉRATION.	
ⲁ ⲃ ⲅ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ	
ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ ⲉ ⲇ ⲉ ⲋ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ	
LETTRES ACCENTUÉES, LIGATURE ET SIGNES DIVERS.	
ⲁ̄ ⲃ̄ ⲅ̄ ⲇ̄ ⲉ̄ ⲋ̄ ⲍ̄ ⲏ̄ ⲑ̄ ⲓ̄ ⲕ̄ ⲗ̄ ⲉ̄ ⲇ̄ ⲉ̄ ⲋ̄ ⲍ̄ ⲏ̄ ⲑ̄ ⲓ̄ ⲕ̄ ⲗ̄	

Figure 4: Coptic Numerals 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).

ⲁ̄ = 1	ⲃ̄ = 20	ⲅ̄ = 300
ⲃ̄ = 2	ⲅ̄ = 30	ⲇ̄ = 400
ⲅ̄ = 3	ⲇ̄ = 40	ⲉ̄ = 500
ⲇ̄ = 4	ⲉ̄ = 50	ⲋ̄ = 600
ⲉ̄ = 5	ⲋ̄ = 60	ⲍ̄ = 700
ⲋ̄ = 6	ⲍ̄ = 70	ⲏ̄ = 800
ⲍ̄ = 7	ⲏ̄ = 80	ⲑ̄ = 900
ⲏ̄ = 8	ⲑ̄ = 90	ⲓ̄ = 1.000
ⲑ̄ = 9	ⲓ̄ = 100	ⲕ̄ = 2.000
ⲓ̄ = 10	ⲕ̄ = 200	ⲗ̄ = 10.000

Figure 5: The representation of numbers in Coptic using letters of the alphabet and horizontal overlines (reproduced from Everson 2003: Figure 12).

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 Coptic Numerals 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: **2009-09-15**
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: **Coptic**
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.)? **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? **No**
 - (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 to write numbers in Coptic.**
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)? **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**