JTC1/SC2/WG2 N4119

L2/09-140 2009-04-11

Doc Type: Working Group Document

Title: Diwani Numerals: Towards a Model for Encoding Numerals of the Siyaq Systems

Source: Anshuman Pandey (pandey@umich.edu)

Status: Individual Contribution **Action:** For consideration by UTC

Date: 2009-04-11

1 Introduction

The intent of this document is to determine possible models for encoding numerals of the Siyaq system in the Universal Character Set (ISO/IEC 10646). It does so through an analysis of the Diwani Numerals, one of the four sub-systems of Siyaq numerical notation.

This document draws upon information originally presented in L2/07-414 "Proposal to Encode Siyaq Numerals in ISO/IEC 10646" (December 2007). In L2/07-414, the present author analyzed the four Siyaq sub-systems and recommended a unified encoding for the numerals of these systems. Although the numerals of the Diwani, Ottoman, Persian, and South Asian traditions are based upon a common typology, there are sufficient differences in character shapes and orthography to warrant an independent encoding for the numerals of each system.

Certain Siyaq traditions have unique requirements for shaping and other rendering behaviors; for example, in the Diwani and South Asian systems there are rules for positioning numerals when writing composite numbers. Certain Siyaq traditions have forms for numerals not found in others; for example, the Persian tradition developed distinct forms for numerals for representing currencies and weights. Also, the Diwani and South Asian systems have alternate forms of the primary numerals that are used for writing composite numerals, while the Ottoman and Persian systems do not. Moreover, certain Siyaq traditions evolved through the influence of local accounting systems; for example, the manner of representing large numerical orders in the South Asian tradition is based not upon the Arabic model, but upon the number system of Sanskrit. Thus, in addition to distinct technical requirements, the four Siyaq traditions differ also on account of their linguistic and historical contexts.

Diwani is the least complex of the four systems of Siyaq. It is, therefore, the system chosen to analyze possible encoding models for Siyaq numerals. It is hoped that a presentation of the typology and numerical notation system of Diwani Numerals will provide information that will facilitate the encoding and implementation of numerals of the Siyaq family in the UCS.

2 Background

The Diwani Numerals are a specialized subset of the Arabic script that were used for maintaining accounting records and other administrative documents. They were developed in the 8th century during the Umayyad caliphate. The numerals originated from the practice of writing numbers using not digits, but the full Arabic names for numbers. As the practice changed through the introduction of abbreviations and calligraphic features, the original Arabic words evolved into distinct monograms. While elements of the original words are visible in a given Diwani numeral, the degree of stylistic innovation masks the relationship between the numerals and the original words. These numerals are not simply presentation forms of the original Arabic letters from which they are derived; they are independent characters that possess particular numerical values.

<i>x</i> 100,000	<i>x</i> 10,000	<i>x</i> 1,000	<i>x</i> 100	<i>x</i> 10	<i>x</i> 1	:	
طالف		الف	6	عا (۶)	1	(1)	1
لاطالف		الفي	Љ	49	ע	(४)	2
سماالف		سالى	الح	てい		(v)	
ليوطالف		لىعالى	لعما	لىع1	لبعا	(لىعر)	4
حطالف		حالف	لوم	1_		()	
رطالف			<i>خ</i> فا	_		(/)	_
بعطالف	-	بعالق	ليم	124	لعا	(L)	7
م. طالف	-	مهالف	لعا	つ	•	(1()	
روطالف	معلا	معالق	لعا	بعا	بعا	(مع)	9

Table 1: Forms of the Diwani Numerals for each order and magnitude

3 The Notation System

Structure Diwani 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.

Directionality Diwani Numerals are written right-to-left in the regular manner of the Arabic script, unlike the left-to-right directionality of the Arabic-Indic digits. The exception is composite numbers of the primary and larger units, which are transposed on account of the manner in which numbers are expressed in Arabic.

Typology Diwani numerals are highly stylized monograms of the Arabic names for numbers for the primary units and their magnitudes in the orders of tens, hundreds, thousands, ten thousands, and hundred thousands. The numerals may be decomposed into basic forms for the numbers 1–10 and distinctive signs that indicate units for different magnitudes (see Section 4 for fuller discussion). The following table illustrates the basic typology with magnitudes of 5 for six decimal orders:

	ر + ع	ح + 1	ح + ط	د + ا + کی) + _	ح + وا + الف
<u>~</u>	لحا	1=	لوے	حالف	حلا	حا الف
BASE	5	50	500	5,000	50,000	500,000

3.1 Ordering

The ordering of Diwani numerals is visual, which reflects the method of expressing numbers in Arabic.

3.2 Orthography

Diwani Numerals are written according to the rules for expressing numbers in Arabic. The largest numeral of a number is written first. The writing of composite numbers is governed by the following rules:

- 1. Composite numbers consisting of the primary numerals and those of the tens, ten thousands, and hundred thousands units are written transposed and with the base form of the primary numeral.
- 2. Composite numbers consisting of the primary numerals and those of the hundreds and thousands units are written using the independent form of the primary numerals in the regular order.
- 3. The numbers 11–19 are written using the base forms of both the primary numeral and TEN.

When written in composite numbers, the base forms of the primary numerals are shaped differently. They are not written fully linearly, but take a cursive shape and extend beneath the following numeral. This shaping feature is shown in the numbers 11–19 in section 4.7.

Examples of the above rules are

- 15 جمسة عشر 'five-ten'. خمسة عشر 'five-ten'.
- 25 عشرون (ع FIVE BASE + علا TWENTY): خمسة و عشرون 'five and twenty'.
- 505 حمل مائة و خمس مائة و خمس مائة و خمس مائة و خمس مائة و خمسة : five hundred and five'. The incorrect form is *حمل مائة و خمس ائة و خمسة : five hundred and five hundred' could theoretically represent 'fifty-five hundred', but this amount would be parsed as 'five thousand and five hundred' and written as حماك حماء.
- 515 حما مائة و خمسة عشر :(five hundred جما FIVE BASE + ۶ TEN BASE) حما معائة و خمسة عشر 'five hundred and five-ten'.
- 50,005 حمسون الفا و خمسة :(fifty thousand and five حلا حا 4 50,005 حلا عا 50,005 •
- 55,000 حمسة و خمسون الفا :(Five and fifty thousand حلا + FIFTY THOUSAND) حمسة و خمسون الفا .
- 55,005 عمسة و خمسون الفا و خمسة (عمله Five Base + عمسة و خمسون الفا و خمسة عمله Five and fifty thousand / and five'.

- 505,505 عالى حالى حالى ما ONE THOUSAND + عالى حالى حالى حالى حالى حالى حالى الله FIVE BASE + والله حالى حالى حالى حالى الله FIVE HUNDRED + حمال أنه الله و خمسة الأف و خمسة الأف و خمسة الأف و خمسة الأف و خمسة أنه و خمسة أنه و خمسة أنه و خمسة أنه و خمسة الأف و خمسة الأ
- 555,555 عا جا FIVE HUNDRED + جا FIVE BASE + عا جا FIFTY THOUSAND + جا جا جا جا FIVE BASE + عا جا FIFTY THOUSAND + جا FIVE BASE + عا FIFTY THOUSAND + خمس مائة و خمسة و خمسة و خمسون الفا و خمس مائة و خمسة و 'five hundred / and five and fifty thousand / and five-hundred / and five and fifty'.

4 The Numerals

4.1 The Primary Unit

The primary unit of Diwani consists of the numbers 1 through 9 and 10. They are stylized monograms of the Arabic names or abbreviations of the names consisting of the initial and one or more letters in a name written with a terminal stroke, which is a stylized representation of the word-final **5** TEH MARBUTA in the names of the units: **1**

1		احد	aḥad	one
ע		اثنان	i <u>t</u> nān	two
ווג		ثلاثة	<u>t</u> alā <u>t</u> a	three
لبعا	لبو+)	اربعة	arba ʻa	four
حا	د + ا	خمسة	<u>h</u> amsa	five
レ	(+ /	ستة	sitta	six
لعا	بعر+)	سبعة	sab ʻa	seven
4	l + v	ثمانية	<u>t</u> amāniya	eight
بعا	رو+ یا	تسعة	tis ʿa	nine
عا	l + s	عشرة	ʿašara	ten

Base Forms of the Primary Numerals The primary numerals may be decomposed to produce base forms:



Variant Forms The following characters have variant forms:

- The base form of 3 (22) takes the shape \searrow when writing tens and hundreds.
- The base form of 8 (\checkmark) takes the shape \checkmark when writing tens and hundreds.

4.2 The Tens Unit

The numerals for 30–90 are composed from the base forms of the primary numerals joined to the tens terminal, which is a stylized form of the $\dot{\upsilon}$ NOON in the Arabic suffix for the tens ($\dot{\upsilon}$), which is represented as a hook: **1**. The exception is 20, which is modeled after \mathfrak{s} , the base form of \mathbf{l} **1**0, as its name $\dot{\upsilon}$ is the dual form of the Arabic name for 10.

49	1 + 9	عشرون	ʻišrūn	twenty
てい	1 + &	ثلاثون	<u>t</u> alā <u>t</u> ūn	thirty
لع1	لىو+ 1	اربعون	arba ʿūn	forty
1=	1+ 🗻	خمسون	<u>h</u> amsūn	fifty
し	1+~	ستون	sittūn	sixty
124	بعر+ 1	سبعون	sab ʿūn	seventy
ں	1+1	ثمانون	<u>t</u> amānūn	eighty
رع1	<i>ر</i> و+ 1	تسعون	tis ʿūn	ninty

4.3 The Hundreds Unit

The numerals for 300–900 are composed from the base forms of the primary numerals joined to the numeral 100 **b**, which is the abbreviation (\Box) of the Arabic word $z\iota$ 'hundred'. The exceptions are 100 **b** and 200 **b**, which are monograms of their Arabic names.

6	_	مائة	mi ʾa	one hundred
Љ	_	مائتان	mi ʾātān	two hundred
الح	P + 🕏	ثلاث مائة	<u>t</u> alā <u>t</u> u miʾa	three hundred
لعبا	لىو+ 1	اربع مائة	arbaʻu mi'a	four hundred
لوم	b + _	خمس مائة	<u>h</u> amsu miʾa	five hundred
خعا	b + ~	ستّ مائة	sittu miʾa	six hundred
لعر	بعر+ لم	سبع مائة	sab ʻu mi 'a	seven hundred
لعا	6 + T	ثمان مائة	<u>t</u> amānu miʾa	eight hundred
پيا	ىع+ ما	تسع مائة	tis 'u mi 'a	nine hundred

Variant Forms The following character has a variant form:

• The numeral THREE HUNDRED (المعلى) also takes the shape المعلى.

4.4 The Thousands Unit

The numerals for 3,000–9,000 are composed from the base forms of the primary numerals joined to the terminal لعن, which is a monogram of the Arabic word الفن 'thousand'. The forms for العن one thousand and two thousand are monograms of their Arabic names.

الف	_	الف	alf	one thousand
العي	_	الفان	alfān	two thousand
سالى	للا + كى	ثلاثة الاف	<u>t</u> alā <u>t</u> a ālāf	three thousand
لىعالى	لبعا + كو	اربعة الاف	arbaʿa ālāf	four thousand
حالف	حا + کی	خمسة الاف	<u>h</u> amsa ālāf	five thousand
ماك	را + ك	ستَّة الاف	sitta ālāf	six thousand
بعالى	بعا + کی	سبعة الاف	sab ʻa ālāf	seven thousand
مهالف	ہا + کی	ثمانية الاف	<u>t</u> amāniya ālāf	eight thousand
معالق	معا + كى	تسعة الاف	tis ʿa ālāf	nine thousand
عالى	عا + كو	عشرة الاف	ʿašara ālāf	ten thousand

Variant Forms The following character has a variant form:

• The numeral ONE THOUSAND (العه) also takes the shape العه).

4.5 The Ten Thousands Unit

The ten thousands are written using modified forms of the tens numerals joined to the terminal **1**, which is a contraction of . The leftward hook in the stylized form of final NOON that marks the tens terminal **1** is dropped and the base is joined to **1**. This is supported by the presence of the variant forms of the base forms of THREE and EIGHT that are used for writing THIRTY and EIGHTY. Moreover, the Arabic names for these numerals supports this typology: 30,000 is 'thirty thousands'. The exception to the typology for the ten thousands is 20,000 **1**, which is modeled after 20 **49**.

lr8	7 +	عشرون الفا	ʻišrūn alfan	twenty thousand
سلا	N +	ثلاثون الفا	<u>t</u> alā <u>t</u> ūn alfan	thirty thousand
لبعلا	N +	اربعون الفا	arbaʿūn alfan	forty thousand
حلا	N +	خمسون الفا	<u>h</u> amsūn alfan	fifty thousand
سلا	N +	ستّون الفا	sittūn alfan	sixty thousand
بعلا	N +	سبعون الفا	sabʻūn alfan	seventy thousand
سلا	N +	ثمانون الفا	<u>t</u> amānūn alfan	eighty thousand
معلا	1 +	تسعون الفا	tis ʿūn alfan	ninty thousand

4.6 The Hundred Thousands Unit

The numerals for the hundred thousands unit are written as U ONE THOUSAND + U ONE HUNDRED + the base form of the primary unit.

طالف	يا + الف	مائة الف	miʾa alf	one hundred thousand
لاماالف	لا + لم + الف	مائتا الف	miʾatā alf	two hundred thousand
يمهاالف	يع + لم + الف	ثلاث مائة الف	<u>t</u> alā <u>t</u> u mi 'a alf	three hundred thousand
لىوطاالف	لىو+ 🖨 + الق	اربع مائة الف	arbaʿu miʾa alf	four hundred thousand
حطالف	ح + لم + الف	خمس مائة الف	<u>h</u> amsu mi 'a alf	five hundred thousand
رطالف	+ له + الف	ست مائة الف	sittu mi 'a alf	six hundred thousand
بعطالف	بعر+ له + الق	سبع ائة الف	sab ʻu mi 'a alf	seven hundred thousand
م. طالف	+ ل + الف	ثمان مائة الف	<u>t</u> amānu miʾa alf	eight hundred thousand
رعواالف	<i>بع</i> + لم + الق	تسع مائة الف	tis ʿu mi ʾa alf	nine hundred thousand

It may be possible to compose the numerals for this unit using #ONE THOUSAND + the hundreds unit of a digit, but this rule is not attested:

4.7 Composite Numbers

The rules for writing composite numbers in the Diwani system are given in Section 3.2. The numbers 10–19 are shown below to illustrate the use of the base form of TEN in writing composite numbers of this range.

عا	_	عشرة	ʿašara	ten
اء	s + 1	احد عشر	aḥad ʿašara	eleven
لاء	۶ + ٦	اثنا عشر	i <u>t</u> nā ʿašara	twelve
کھ	۶ + 0	ثلاثة عشر	<u>t</u> alā <u>t</u> a ʿašara	thirteen
لنوع	لبو+ ء	اربعة عشر	arbaʿa ʿašara	fourteen
82	۶ + 🇻	خمسة عشر	<u>h</u> amsa ʿašara	fifteen
9	9+~	ستة عشر	sitta ʿašara	sixteen
بعي	بعر+ ء	سبعة عشر	sabʻa ʻašara	seventeen
97	9 + T	ثمانية عشر	<u>t</u> amāniya ʿašara	eighteen
بعي	ىع+ ب	تسعة عشر	tis ʿa ʿašara	nineteen

Composite numbers from 21–99 are also written with the base form of the primary numeral and the respective tens numeral. They are expressed using the conjunction wa 'and'. Thus, 21 is written wa and is expressed as 'expressed' one and twenty', 22 is wa 'two and twenty', etc.

5 Implementation

5.1 Encoding Model

Given the above analysis, the possible models for encoding the Diwani Numerals are:

- 1. Encode each numeral as an atomic character
- 2. Encode the base forms of the primary units and unit marks
- 3. Encode the numerals for the primary, tens, hundreds, thousands, and ten thousands units
- **1. Encode each numeral as an atomic character** The most elementary approach to encoding the Diwani Numerals is to encode each individual numeral as an atomic character. This model would require 69 characters: primary units (10), base forms of the primary units (10), tens (9), hundreds (10), thousands (10), ten thousands (10), and hundred thousands (10).

The advantage of this model is that no special rendering rules are needed to write the numerals. The disadvantage is the encoding of redundant characters, in particular the hundred thousands unit, which may be written using characters for other units.

2. Encode the base forms of the primary units and unit marks This is an extreme alternative to encoding each numeral as an atomic character. It is a means of encoding Diwani Numerals according to their typological decomposition. In this approach, the Diwani Numerals would be written using the base forms of the primary numerals and the distinctive sign for each decimal order. This approach would require only 16 characters: base forms of the primary units (10) and signs for the units (6).

The major disadvantage to this approach is the heavy reliance upon rendering rules. The shaping engine would need to produce the appropriate forms for special ligatures. The number one thousand would be produced using thousands sign + ∞ one; ten thousand would be ∞ ten thousands sign + ∞ ten; twenty thousand be would be ∞ ten thousands sign + ∞ ten thousands sign + ∞ two.

Another disadvantage is ordering. With this approach the rendering engine would need to first compose the appropriate number for a base numeral + a unit sign, then order these pairs according to the Arabic counting order.

Although the primitives approach reflects the pattern that underlies the typology of the Diwani Numerals, the complexity of this encoding model may restrict its implementation.

3. Encode the numerals for the primary through ten thousands units A third approach is a mean between the two discussed previously. In this approach, the numerals of the primary, tens, hundreds, thousand, and ten thousands units are encoded as atomic characters. Based upon their glyphic representation, the numerals for these units are unique and cannot be represented using other characters (apart from the primitives

model). The numerals for the hundred thousands unit may be written using the base forms of the primary unit + ONE HUNDRED + ONE THOUSAND.

This model would require 59 characters: primary units (10), base forms of the primary units (10), tens (9), hundreds (10), thousands (10), and ten thousands (10). Of the three, this approach offers the least complicated method of encoding Diwani Numerals.

5.2 A Basic Character Set for Diwani Numerals

Based upon encoding model #3, 59 characters are required to encode Diwani Numerals in the UCS:

```
xx01 DIWANI NUMERAL ONE
xx02 DIWANI NUMERAL TWO
xx03 DIWANI NUMERAL THREE
xx04 DIWANI NUMERAL FOUR
xx05 DIWANI NUMERAL FIVE
xx06 DIWANI NUMERAL SIX
xx07 DIWANI NUMERAL SEVEN
xx08 DIWANT NUMERAL EIGHT
xx09 DIWANI NUMERAL NINE
xx0A DIWANI NUMERAL TEN
xx0B DIWANI NUMERAL COMBINING ONE
xx0C DIWANI NUMERAL COMBINING TWO
xx0D DIWANI NUMERAL COMBINING THREE
xx0E DIWANI NUMERAL COMBINING FOUR
xx0F DIWANI NUMERAL COMBINING FIVE
xx10 DIWANI NUMERAL COMBINING SIX
xx11 DIWANI NUMERAL COMBINING SEVEN
xx12 DIWANI NUMERAL COMBINING EIGHT
xx13 DIWANI NUMERAL COMBINING NINE
xx14 DIWANI NUMERAL COMBINING TEN
xx15 DIWANI NUMERAL TWENTY
xx16 DIWANI NUMERAL THIRTY
xx17 DIWANI NUMERAL FORTY
xx18 DIWANI NUMERAL FIFTY
xx19 DIWANT NUMERAL SIXTY
xx1A DIWANI NUMERAL SEVENTY
xx1B DIWANI NUMERAL EIGHTY
xx1C DIWANI NUMERAL NINETY
xx1D DIWANI NUMERAL ONE HUNDRED
xx1E DIWANI NUMERAL TWO HUNDRED
xx1F DIWANI NUMERAL THREE HUNDRED
xx20 DIWANI NUMERAL FOUR HUNDRED
xx21 DIWANI NUMERAL FIVE HUNDRED
XX22 DIWANI NUMERAL SIX HUNDRED
XX23 DIWANI NUMERAL SEVEN HUNDRED
xx24 DIWANI NUMERAL EIGHT HUNDRED
xx25 DIWANI NUMERAL NINE HUNDRED
xx26 DIWANI NUMERAL ONE THOUSAND
xx27 DIWANI NUMERAL TWO THOUSAND
xx28 DIWANI NUMERAL THREE THOUSAND
xx29 DIWANI NUMERAL FOUR THOUSAND
xx2A DIWANI NUMERAL FIVE THOUSAND
xx2B DIWANI NUMERAL SIX THOUSAND
xx2C DIWANT NUMERAL SEVEN THOUSAND
xx2D DIWANI NUMERAL EIGHT THOUSAND
xx2E DIWANI NUMERAL NINE THOUSAND
xx2F DIWANI NUMERAL TEN THOUSAND
xx30 DIWANI NUMERAL TWENTY THOUSAND
xx31 DIWANI NUMERAL THIRTY THOUSAND
xx32 DIWANI NUMERAL FORTY THOUSAND
xx33 DIWANI NUMERAL FIFTY THOUSAND
xx34 DIWANI NUMERAL SIXTY THOUSAND
xx35 DIWANI NUMERAL SEVENTY THOUSAND
xx36 DIWANI NUMERAL EIGHTY THOUSAND
xx37 DIWANI NUMERAL NINETY THOUSAND
```

6 References

Kazem-Zadeh, H. 1915. "Les Chiffres Siyâk et la Comptabilité Persane." In *Revue du Monde Musulman*, vol. 30, pp. 1–51.

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

LES CHIFFRES « DÎVÂNÎ » CHEZ LES ARABES (I)

CHIFFRES	VALEUR	CHIFFRES	VALEUR	CHIFFRES	VALEUR
6	I	يعدو	19	الف ou للبه	1,000
y	2	25	20	العي	2,000
ம ₀u W	3	س	30	ساك	3,000
لعا	4	1 eJ	40	لعالف	4,000
ما	5	1ء	50	حالف	5,000
L	6	レ	60	ماهد	6,000
اما	7	121	70	معالف	7,000
4	8	ا ل	80	مهاکف	8,000
لعا	9	لعا	90	بعاك	9,000
عا	10	6	100	عالف	10,000
١ء	11	љ	200	ومنا	20,000
لاع	I 2	ملیا ۵۰۰	300	سلا	30,000
جع	13	لمعا	400	لحلا	40,000
لعء	14	حما	500	حلا	50,000
حري	15	لفا	600	y_	60,000
رع	16	Les	700	معلا	70,000
معو	17	لما	800	سلا	80,000
5-V	18	بعا	900	كلا	90,000

⁽¹⁾ D'après un manuscrit du Vocabulaire arabe-persan de Zamakhchari (Bibliothèque Nationale, ancien fonds arabe nº 1256), reproduits dans la Grammaire arabe de Silvestre de Sacy et dans l'ouvrage de A.-P. Pihan.

Figure 1: Table showing the Diwani number forms (from Kazem-Zadeh 1915: Plate VII).

unités.		DIZAIN	ES.	CENTAINES.	
1	1	عا	10	6	100
ע	9	49	20	.No	200
or on m≀	3	てい	3о	الموس ۱۵۰ لك	300
لبعا	4	لبع1	40	ليها	400
حا	5	15	5o	لعم	500
レ	6	レ	6o	خعا	600
لعا	7	124	70	لعا	700
4	8	つ	80	لعا	800
لعا	9	رع1	90	لعا	900
MILLE	MILLE.		DIZAINES DE MILLE.		E MILLE.
الف ٥١ الك	1,000	عالی	10,000	طالف	100,000
العي	2,000	lh8	20,000	لاطالف	200,000
سالف	3,000	ىىلا	30,000	يمطالف	300,000
لىعالى	4,000	لبعلا	40,000	لبعرطالف	400,000
حالف	5,000	حلا	50,000		
راك	6,000	سلا	60,000		
بعالف	7,000	بعلا	70,000		
مهای	8,000	ىلا	. 80,000		
تعالق	9,000	معلا	90,000		

Figure 2: Table showing the Diwani number forms (from Pihan 1860: 211).

EXEMPLES DE QUELQUES NOMBRES COMPOSÉS.							
اء	11	بعء	17	لالبع1	42		
لاء	12	5-1	18	العلى	48		
500	13	بعي	19	طالبعا	141		
لبوع	14	rel	21	مالاحا	152		
92	15	لبعرون	24	→ S6	206		
9	16	حرب 1	35	ملحاحء	315		

Figure 3: Table showing composite numbers written with Diwani Numerals (from Pihan 1860: 212).