| Doc Type: | Working Group Document |
| :--- | :--- |
| Title: | Raqm Numerals: A Model for Encoding the Siyaq System of South Asia |
| Source: | Anshuman Pandey (pandey@umich.edu) |
| Status: | Individual Contribution |
| Action: | For consideration by UTC |
| Date: | $2009-04-21$ |

## 1 Introduction

The intent of this document is to determine models for encoding numerals of the Siyaq notation system in the Universal Character Set (ISO/IEC 10646). It does so through an analysis of the Raqm Numerals, the Siyaq sub-system used in South Asia.

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 traditions and recommended a unified encoding for the numerals of these sub-systems. Further research has indicated that although the numerals of the Diwani, Ottoman, Persian, and South Asian traditions are based upon a common pattern, there are sufficient differences in character typology and orthography to warrant the independent encoding for the numerals of each sub-system.

This document is intended to stand in comparison to L2/09-140 "Diwani Numerals: Towards a Model for Encoding Numerals of the Siyaq Systems" (April 2009). In L2/09-140, the present author described the Diwani Numerals, which possesses the smallest character repertoire of the four sub-systems and the least technical requirements for shaping and other rendering behaviors. The presentation here of the Raqm system may be compared with that of the Diwani Numerals in order to understand not only the differences in the character typology of the numerals of the Siyaq family, but also the locale-specific linguistic factors that differentiates the Siyaq sub-systems.

## 2 Background

In South Asia, the Siyaq system is known as Raqm (Arabic رقم raqm 'account'). Similar to the other Siyaq traditions, the Raqm Numerals are a specialized subset of the Arabic script that was used for accounting and other numerical notation. The basic Raqm Numerals are stylized monograms of the Arabic names for the numbers, but the numerals for large decimal orders are derived not from Arabic, but from Indic languages, and the method of writing fractions and currencies is based on a common north Indic numerical notation system.

While the majority of documents containing Siyaq are hand-written, a rare exception is the appearance of Raqm Numerals in printed books. A work by Francis Gladwin titled $A$ Compendious System of Bengal Revenue Accounts (1790) is perhaps the first book to use Siyaq-based numerals in print. In the preface to his book, Gladwin writes "that the following compendium of Siyak Accounts is the first specimen of the sort that has yet appeared in print, the types having been made purposely for it." ${ }^{1}$ A specimen of Raqm printed in Gladwin's metal fonts is given in Figure 9.

[^0](

Table 1: South Asian forms of the Siyaq numerals for six decimal orders.

## 3 The Notation System

### 3.1 Structure

Raqm 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.

### 3.2 Directionality

Raqm 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.

### 3.3 Typology

Raqm 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 and ten thousands. Raqm Numerals for the hundred thousands and millions are based upon names for numbers derived from Sanskrit.

The numerals may be decomposed to some degree into basic forms and into distinctive signs for the various decimal orders. The exceptions are generally the numerals for magnitudes of one ( $1,10,100$, etc.) and two ( $2,20,200$, etc.), which have unique forms in all decimal orders. A complete description of the numerals of each decimal order is given in section 4 , however, a summary is given below:

- The independent shape of a primary numeral can be decomposed into a character primitive that forms

- Each primary numeral has a base form that is used in writing composite numbers. The base form of a primary numeral is produced by joining the primitive form of the primary numeral to a horizontal stroke: $\boldsymbol{\sim}$ FIVE baSE $\leftarrow \boldsymbol{\sim}$ base mark $+\boldsymbol{\omega}$ Five primitive.
- The numerals for the tens consist of the base form of the primary numeral joined to a distinctive

- The numerals for the hundreds consist of the base form of the primary numeral joined to the numeral for 100: $\boldsymbol{\bullet}_{\text {FIVE HUNDRED }} \leftarrow \mathfrak{C}$ one hundred $+\boldsymbol{\omega}_{\text {FIVE BASE. Certain numerals for the hundreds use }}$ an alternate base form.
- The numerals for the thousands consist of the primitive form of the primary numeral joined to a distinctive terminal: صم five thousand $\leftarrow \sim$ thousands mark $+\boldsymbol{\sim}$ FIVe primitive. When written independently, the thousands are marked with the placeholder $\overline{\bar{*}}$, eg. $\xrightarrow{\bar{\sim}}$ FIVE THOUSAND
- The numerals for the ten thousands consist of a modified (base) form of the tens numerals, which is produced by dropping the $\mathbf{Q}$ TENS MARKS, and a placeholder when written independently: $\boldsymbol{\Omega}_{\text {FIFTY }}$ THOUSAND $\leftarrow{ }^{〔}$ PLACEHOLDER $+\sim$ FIFTY bASE. For most tens numeral, the base is identical to the base form of the primary numeral, but different for TEN, TWENTY, and EIGHTY, eg. the base form of EIGHTY is and the base form of eight is
- The numerals for the hundred thousands are written using the independent form of the primary numeral and a mark that represents the hundred thousands unit: $\quad$ FIVE HUNDRED THOUSAND $\leftarrow$ HUNDRED THOUSANDS MARK $+{ }^{-}$FIVE
- The numerals for the millions are written using a modified (base) form of the tens (similar to the
 hUNDRED THOUSANDS MARK + FIFTY baSE
- The numerals for the ten millions are written using the independent form of the primary numeral and



### 3.4 Ordering

The ordering of Raqm Numerals is visual, which reflects the method of expressing numbers in Arabic. The ordering of Raqm Numerals based upon non-Arabic sources is also visual, which reflects the method of expressing numbers in Urdu.

### 3.5 Orthography

Raqm Numerals are written according to the rules for expressing numbers in Arabic. The largest numeral of a number is written first, except in composite numbers.

The writing of composite numbers (primary numeral + larger numeral) is governed by the following rules:

1. The base form of the primary numeral is used in composite numbers of the same decimal order. The numerals are written transposed, with the primary numeral positioned before the larger numeral. This rule governs the writing of numbers such as $10,11, \ldots, 10,000,11,000, \ldots$, etc.
2. The independent form of the primary numeral is used in composite numbers of different decimal orders. The numerals are written in the regular order with the larger numeral preceding the primary numeral. This rule governs the writing of numbers such as $101,102, \ldots, 1,001,1,002, \ldots, 1,100,1,200$, ..., etc.

## 4 The Numerals

### 4.1 The Primary Unit

The primary unit consists of numerals for the numbers 1 through 9. The numerals are stylized monograms of the Arabic names for the numbers or abbreviations of the names consisting of the initial and one or more letters. The glyphs for Raqm one and Two have origins different from those of the numerals in other Siyaq sub-systems, where the numeral one is derived from احدا ahad 'one' and Two from اثنان itnān 'two'. The Raqm numeral ONE is based upon Arabic عدد 'dad 'number' and Two is based upon عددن 'dadan 'dual'.

| RAQM | COMPOSITION | ARABIC SOURCE |  | ENGLISH |
| :---: | :---: | :---: | :---: | :---: |
| ع | $0+2+\varepsilon$ | علد | 'dad | one |
| Cos | $j+2+\varepsilon$ | عددان | 'dadān | two |
| $\mu$ | $\mathcal{L}+J+ \pm$ | ثلاثة | talāta | three |
| $b^{y}$ | $\varepsilon+\jmath+1$ | اربعة | arba 'a | four |
| صم | ${ }^{+}$ | خمسة | hamsa | five |
| $\simeq$ | ${ }^{\sim}+\cdots$ | ستّة | sitta | six |
| معه | $0+$ + | سبعة | $s a b ' a$ | seven |
| $\underline{L}$ | ${ }^{\sim}+1+\sim$ | ثمانية | $\underline{\text { tamāniya }}$ | eight |
| 0 | $\varepsilon+ت$ | تسعة | tis ' $a$ | nine |

### 4.1.1 Variant Forms

The following characters have variant forms:

- The numeral one is also written as عص
- The numeral عهع Two is also written as $C$ and $C$.


### 4.1.2 Base Forms of the Primary Numerals

When written in composite numbers, the primary numbers are written differently, in what might be called their base form. These base forms are derived from the most primitive element of the numeral. This character primitive is joined to a horizontal swash (comparable to _ U+0640 ARABIC TATWEEL) to create the base form. The character primitive is used in the numerals for the hundreds and thousands, with some exceptions as described in the sections for those units.


### 4.2 The Tens Unit

The numerals for the tens unit are composed from the base forms of the primary numerals joined to the tens terminal, which is a stylized form of the $\dot{U}$ NOON in the Arabic suffix for the tens ( $\dot{U} \boldsymbol{g} \bar{u} n)$, which is represented as a loop: ©. The exceptions are TEN and TWENTY, which have base forms that are used in

 context.

| RAQM | COMPOSITION | ARABIC SOURCE |  | ENGLISH |
| :---: | :---: | :---: | :---: | :---: |
| عه0 | c. | عشرة | 'ašara | ten |
| عسه | c. | عشرون | 'išrūn | twenty |
| Or | c. | ثالاثون | $\underline{\text { talāt̄}}$ | thirty |
| للمــــهـ | C. | اربعون | arba 'ūn | forty |
| - | $0+$ - | خمسون | $\underline{\text { hamsūn }}$ | fifty |
| 0 | C+ | ستّون | sittūn | sixty |
| هوـ0 | م | سبعون | $s a b$ 'un | seventy |
| Q | $0+\square$ | ثمانون |  | eighty |
| - | Q | تسعون | tis ' ${ }^{\text {¢ }}$ n | ninty |

### 4.2.1 Base Forms of the Tens Numerals

The tens numerals have base forms that are used in the writing of numbers of the ten thousands and millions units. The base form of a tens numeral is derived by dropping the stylized nOON that marks the tens terminal a. Except for TEN, TWENTY, and EIGHTY, the base forms of the tens are identical to the base form for the corresponding primary unit.

|  | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| INDEPENDENT | Q | هm | 0 | 0 | 0 |  | 0 | 0 | - |
| BASE | ع | 1 |  |  |  |  | - |  | , |

### 4.3 The Hundreds Unit

The numerals for the hundreds are derived from the primitive forms of the primary numerals joined to the mark 6 , which is the abbreviation of the Arabic word مائة 'hundred' as 0 . However, not all primitive forms of the primary numerals used in writing the hundreds are identical to those described in section 4.1.2. The primitives used in writing the numerals for FOUR HUNDRED, SIX HUNDRED, SEVEN HUNDRED, and EIGHT HUNDRED represent a different mutation of the primary numerals. These mutations are described below. Apart of these, the other exceptions are the numerals ONE HUNDRED and TWO HUNDRED, which are monograms of the Arabic names for these numbers.

| RAQM | OMPOSITION | ara | URCE | ENGLISH |
| :---: | :---: | :---: | :---: | :---: |
| 6 | - | مائة | mi'a | one hundred |
| $\Omega$ | - | مائتّان | mi'ātān | two hundred |
| 6 | $6+\sim$ | ثلاث مائة | tılātu mi'a | three hundred |
| 6 | $6+\sim$ | اربع مائة | arba'u mi'a | four hundred |
| 6 | $6+\infty$ | خمس مائة | $\underline{\text { hamsu mi'a }}$ | five hundred |
| 6 | $6+\sim$ | ستّ مائة | sittu mi'a | six hundred |
| 4 | $6+1$ | سبع مائة | sab ' mi'a $^{\text {a }}$ | seven hundred |
| 6 | $6+J$ | ثمان مائة | tıamānu mi'a | eight hundred |
| $w$ | $6+9$ | تسع مائة | tis 'u mi'a | nine hundred |

The following numerals take forms different from the base shape when used to form the hundred numeral

|  | HUNDREDS |  | PRIMITIVE | Independent |
| :---: | :---: | :---: | :---: | :---: |
| 400 | 6 | $\cdots$ | للو | 5 |
| 600 | G | $\sim$ | $\sim$ | $\simeq$ |
| 700 | 4 | 1 | - | م |
| 800 | 6 | $\checkmark$ | $\checkmark$ | $\underline{L}$ |

## 4．4 The Thousands Unit

The numerals for the thousands are composed from the primitive forms of the primary numerals joined to the terminal $\quad$ ，which is an abbreviation of the Arabic word الف＇thousand＇．When written independently， the thousands are marked as ${ }^{\bar{Z}}$ ，where the element ${ }^{\overline{ }}$ represents the absence of other numbers．
It may appear that the thousands are written using the base form of the primary numeral + the element ${ }^{\overline{ }}$ ， but this is not the case．Only the two thousand follows this pattern：cf．base form of two and two ThOUSAND E．However，this is on account of the special forms for two in each decimal order．Compare base form of six－and six thousand Er．

The forms for one thousand and two thousand have special forms．The one thousand is a monogram


| RAQM | COMPOSITION | ARABIC | OURCE | ENGLISH |
| :---: | :---: | :---: | :---: | :---: |
| الـ］ | － | الف | alf | one thousand |
| 5 | － | الفان | alfān | two thousand |
| تمـ0r | $\xrightarrow{\square}+\sim$ | ثالاثة الاف | țalāta ālāf | three thousand |
| \＃3 |  | اربعة الاف | arba＇a ālāf | four thousand |
| صمـ | $\underline{\square}+\infty$ | خمسة الاف | hamsa àlāf | five thousand |
| Er | $\underset{\sim}{\square}+\sim$ | ستّة الاف | sitta ālāf | six thousand |
| معـ三丨三丨 | $\underline{\sim}$ | سبعة الاف | $s a b$＇a ālāf | seven thousand |
| EN | $\underset{\sim}{\underline{\underline{H}}+\sim}$ | ثمانية الاف | tıamāniya ālāf | eight thousand |
| لمـــ三 | $\xrightarrow{\square}$ | تسعة الاف | tis＇a ālăf | nine thousand |

## 4．4．1 Variant Forms

The following characters have variant forms：
－The numeral الـــة one thousand also takes the shape السَّ
－The numeral שس Two thousand also takes the shape

## 4．5 The Ten Thousands Unit

The ten thousands are written using modified forms of the tens numerals，in which the stylized noon that marks the tens terminal $\boldsymbol{Q}_{\text {－}}$ is dropped．The base forms used in writing the ten thousands unit are derived from modified forms of the tens and are not the base forms of the primary numerals．This assertion is supported by the original Arabic names for these numbers，eg．ثمانون الفا tamānūn alfan＇eighty thousands＇．Furthermore，the base forms of ten，twenty，and eighty are used in writing these numerals， eg． $80,000 \underbrace{\sim}$ and not＊When written independently，the ten thousands are marked with＂ ，which is a placeholder mark that represents the absence of other numbers．

| RAQM | COMPOSITION | ARABIC SOURCE |  | ENGLISH |
| :---: | :---: | :---: | :---: | :---: |
| عه\% | $\cdots+$ c | عشرة الاف | 'ašara ālāf | ten thousand |
| عس" | cm+ | عشرون الفا | 'išrūn alfan | twenty thousand |
| N+ | $\cdots+$ | ثلاثون الفا | ṫalātūn alfan | thirty thousand |
| لهس | " | اربعون الفا | arba 'ūn alfan | forty thousand |
| + | $\cdots+$ | خمسون الفا | $\underline{h} a m s u ̄ n ~ a l f a n ~$ | fifty thousand |
|  | " + | ستّون الفا | sittūn alfan | sixty thousand |
|  | "m+ | سبعون الفا | sab 'ūn alfan | seventy thousand |
|  | $\cdots+\cdots$ | ثمانون الفا | tamānūn alfan | eighty thousand |
| ****) | ") + | تسعون الفا | tis 'ūn alfan | ninty thousand |

### 4.5.1 Variant Forms

The following characters have variant forms:

- The numeral $\underbrace{\text { Cen thousand also takes the shape }}_{\text {ת }}$, which is based upon the typology of the thousands. It is derived from the word-initial AIN of عشرة 'ašara 'ten' $+\rightarrow$ the thousands base. In some Siyaq sub-systems, TEN is grouped with the primary units and, therefore, its numerals of the various decimal orders are based upon the typology for that group.


### 4.6 The Hundred Thousands Unit

The numerals for the hundred thousands unit are written as the regular form of the primary unit $+\boldsymbol{J}$ hundred thousands mark. This method of representing this decimal order in the Raqm tradition differs from the other Siyaq sub-systems in that it borrows from a non-Arabic tradition. The monogram $\boldsymbol{J}$ is derived from the Hindi लाख 'hundred thousand'. In modern Indian notation, the lākh is written as 1,00,000.

This is different from the Arabic model, where 'hundred thousand' is expressed as مائة الف mi'a alf and various magnitudes of the unit are expressed by prefixing the primary numeral to the unit, eg. 'five hundred thousand' خمس مائة الف hamsu mi'a alf.

The Raqm tradition reflects the typology of the Siyaq system in that it has special forms for ONE HUNDRED THOUSAND and Two HUNDRED THOUSAND; these are written as

RAQM COMPOSITION URDU SOURCE ENGLISH

| N/NTM |  | اكي لاكه | $\bar{e} k l a \bar{k} h$ | one hundred thousand |
| :---: | :---: | :---: | :---: | :---: |
| $0 \cos$ |  | ， | dō lākh | two hundred thousand |
| 促 | U | تّ． | tīn lākh | three hundred thousand |
| Nهك | ك＋ | چار لاكه | chār lākh | four hundred thousand |
| مك | هم＋ | 比 | pānch lākh | five hundred thousand |
| \％ |  | چ | chah lākh | six hundred thousand |
| هوس | $\checkmark+$ | ¢ا لا | sāt lākh | seven hundred thousand |
| U年 | U | آّ | āth lākh | eight hundred thousand |
| وهك | ك | ¢ | $n o ̄ l a ̄ k h$ | housan |

The writing of the hundred thousands unit reflects the expression of numbers of the group．The number
 not as＊（THREE bASE＋HUNDRED THOUSANDS MARK）．

## 4．7 The Millions Unit

The numerals for the millions unit are written with the base form of the tens and $\boldsymbol{\sim}$ HUNDRED thousands мARK．The millions are an extension of the hundred thousands unit and are expressed as＇tens of hundred－ thousands＇，eg．five million is＇fifty hundred－thousands＇．This system is unrelated to the Arabic model， where the millions unit，in the Ottoman Siyaq tradition，is expressed as＇thousand times a thousand＇，eg．five million خمسة الاف مرة الف hamsu a aläf maratan alf＇five－thousand times a thousand＇．In modern Indian notation，the millions are expressed as＇ten lākh＇and written as $10,00,000$ ．
COMPOSITION $\quad$ URDU SOURCE $\quad$ ENGLISH

### 4.8 The Ten Millions Unit

The numbers for the ten millions are an extension of the millions - 'hundreds of hundred-thousands' or 'one-hundred lākh'. In modern Indian notation, the tens millions are expressed as karor ('hundred lākh') and written as $1,00,00,000$.

The numeral ten million is $\boldsymbol{\beta}$, which is derived from karor (Hindi करोड़ karor $<$ Sanskrit कोटि koti). The numeral twenty million is written אرתا karorān, which is expressed as a plural of that adds the Persian plural marker كرور $\begin{aligned} \text {. } \\ \text { آن } \\ \text {. This illustrates an attempt to pattern non-Arabic number names on an }\end{aligned}$ Arabic pattern, similar to the method in which 10 and 20 are expressed in Arabic as عشرون 'ašara and ' 'išrūn.

| RAQM | COMPOSITION | URDU SOURCE | ENGLISH |
| :---: | :---: | :---: | :---: |
|  | - | karōr | ten million |
| - |  | karōrān | twenty million |

The numbers $30-90$ million are written using the primary numeral + the numeral $\boldsymbol{\beta}_{\mathrm{J}}^{\mathrm{TEN}}$ million, eg. 50 million is 'five karor' and is written صم كגر.

Hypothetically, the method of representing 100 millions would be patterned upon the typology for the ten millions. The number 500 million would be rendered as 'fifty karor'' and written as مسه תנر.

### 4.9 Composite Numbers

Composite numbers of the primary and tens units are written using the base form of the primary numeral and the appropriate tens numeral. The numbers 10-19 are illustrated below. In some Siyaq sub-systems, such
as Diwani, the numbers 10-19 are written using a base form of ten. Although the Raqm system has a base form of TEN, all composite numbers are written using the regular forms of the tens numerals.

| RAQM | Composition | arabic source |  | English |
| :---: | :---: | :---: | :---: | :---: |
| عـ0 | - | عشرة | 'ašara | ten |
| 1 | cel | احد عشر | ahad 'ašara | eleven |
| תعـع | عسـ + ع | اثنا عشر | itinā 'ašara | twelve |
| c | $0 \times+$ | ثلاثة عشر | talāta 'ašara | thirteen |
| Nתع | للهـ + عـهـ + | اربعة عشر | arba 'a 'ašara | fourteen |
| قعك |  | خمسة عشر | hamsa 'ašara | fifteen |
| C | ع | ستّة عشر | sitta 'ašara | sixteen |
| תوعه | - ع- + | سبعة عشر | sab 'a 'ašara | seventeen |
|  | - + | ثمانية عشر | tıamāniya 'ašara | eighteen |
| لوعه | لـهـ + عـه | تسعة عشر | tis 'a 'ašara | nineteen |

As the table above illustrates, the base form of the primary numeral is positioned in a particular fashion with the larger numeral. If the base form of the primary numeral has no vertical element on the right edge, it is written beneath the larger numeral: $16 \leftarrow$ שعـ TEN + SIX BASE. If a vertical element is present, the swash of the primary numeral subtends beneath the numeral: عـه TEN 19 لوعس TENE BASE.

When the thousands and ten thousands are written with numerals from smaller decimal orders, the place-



### 4.10 Fractions

There are three signs for representing fractions in the Raqm system.

| RAQM |  | ENGLISH |
| ---: | :---: | :--- |
| - | $1 / 4$ | one quarter |
| • | $1 / 2$ | one half |
| - | $3 / 4$ | three quarters |

### 4.11 Placeholder Mark

As described in sections 4.4 and 4.5, the thousands and ten thousands are written when the placeholder mark = when they appear alone. This mark is also written as "" and $\boldsymbol{w}^{\prime}$. This mark represents the absence of numbers and is written above the horizontal stroke of numerals: ${ }^{\bar{Z}}$.

### 4.12 Currency Sign

The character / is used to write currencies. It represents the rupee currency unit. There is a special orthography for writing currencies in Raqm, which is based on the rupee system that was common across northern South Asia (see Pandey 2007a for a description). This currency system is based upon the rupee (rupayā), $\bar{a} n \bar{a}$, and $p \bar{a} \bar{\imath}$; with regional variants such as that used in Bengali, which uses the unit gand $\bar{a}$ instead of $p \bar{a} \bar{l}$, as shown in Figure 7 (see Pandey 2007b for a description).

When a Raqm numeral is written with the currency mark, it represents rupees: 'fifty rupees' مسـه/. When an Arabic-Indic digit is written with a currency mark, it represents values of the $\bar{a} n \bar{a}$ unit: 'five $\bar{a} n \bar{a}$ ’ $\boldsymbol{\Delta}$. When a fraction sign is written with the currency mark, the combination represents a fraction of the $\bar{a} n \bar{a}$ unit, which is called the $p \bar{a} \bar{l}$ unit: ' $1 / 4 \bar{a} n \bar{a}$ ' or 'one $p \bar{a} \bar{l}$ ' $/{ }^{\prime}$.

## 5 Implementation

### 5.1 Encoding Model

Given the above analysis, there are three possible model for encoding the Raqm Numerals.

1. Encode each numeral as an atomic character
2. Encode the numerals using character primitives
3. Encode a combination of numerals and unit marks
4. Encode each numeral as an atomic character The most elementary approach to encoding the Raqm Numerals is to encode each individual numeral as an atomic character. This model would require 85 characters for the numerals: primary units (9), base forms of the primary units (9), tens (9), hundreds (9), thousands (9), ten thousands (9), hundred thousands (9), and millions (9), and ten millions (9); and fractions (3) and currency mark (1).

The advantage of this model is that no special rendering rules are needed to write the numerals. Units larger than millions may be written using combinations of other characters.

The disadvantage is the encoding of redundant characters, in particular the hundred thousands, millions, and ten millions units, which may be written using characters for other units.
2. Encode the numerals using character primitives While the typological characteristics of the Diwani Numerals makes it theoretically possible to encode that Siyaq sub-system using character primitives, this model does not provide an effective means for encoding the Raqm Numerals. In this approach, the Raqm Numerals would be represented using the primitive forms of the primary numerals and the distinctive sign for each decimal order, as is a possibility for Diwani Numerals. However, while there are distinctive signs for tens and other units, there is no such sign for the primary units in Raqm.

To encode Raqm numerals using character primitives, a *PRIMARY UNITS MARK would have to be invented. Raqm one would then be produced by writing *Primary units mark $+\lambda$ one base, requiring the rendering engine to produce the correct glyph from the backstore from a given a sequence of characters.

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 $\equiv$ would be produced by Thousands SIGN $+\boldsymbol{\downarrow}$ ONE base.

As the thousands and ten thousands are not distinguished through terminal marks, but by the base shape of the numeral, producing numerals for ten thousands would be expensive in a character-primitives model. The
 Placeholder mark $+\boldsymbol{\sim}$ thousands mark $+\boldsymbol{\sim}$ TENS mark $+\boldsymbol{\jmath}_{\text {base one; }}$

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 Raqm Numerals, the complexity of this encoding model will restrict its implementation.
3. Encode a combination of numerals and unit marks A third approach is a mean between the two discussed previously. In this model the numerals of the primary, tens, hundreds, thousands, and ten thousands units are encoded as atomic characters. Based upon their glyphic representation, the various numerals for the hundred thousands and millions may be written using sequences of other characters, eg. the numerals for the hundred thousands unit may be written using the primary unit + one hundred thousands mark. This model also encodes numerals of various decimal orders, which may be considered typologically unique.

This model would require 64 characters:

- The primary numerals and their combining forms (18)
- The tens (9)
- The hundreds (9)
- The thousands (9)
- The ten thousands (9)
- ONE HUNDRED THOUSAND, TWO HUNDRED THOUSAND, and HUNDRED THOUSANDS MARK (3)
- one million and two million (2)
- Fraction signs (3)
- Currency marks (1)
- Placeholder mark (1)

Of the three, this approach offers the least complicated method of encoding Raqm Numerals.

### 5.2 A Basic Character Set for Raqm Numerals

Based upon encoding model \#3, 64 characters are required to encode Raqm Numerals in the UCS:

```
xx01 RAQM NUMERAL ONE
xx02 RAQM NUMERAL TWO
xx03 RAQM NUMERAL THREE
xx04 RAQM NUMERAL FOUR
xx05 RAQM NUMERAL FIVE
xx06 RAQM NUMERAL SIX
xx07 RAQM NUMERAL SEVEN
xx08 RAQM NUMERAL EIGHT
Xx09 RAQM NUMERAL NINE
xxOA RAQM NUMERAL TEN
xxOB RAQM NUMERAL TWENTY
xxOC RAQM NUMERAL THIRTY
xxOD RAQM NUMERAL FORTY
xx0E RAQM NUMERAL FIFTY
xxOF RAQM NUMERAL SIXTY
xx10 RAQM NUMERAL SEVENTY
xx11 RAQM NUMERAL EIGHTY
xx12 RAQM NUMERAL NINETY
```

```
xx13 RAQM NUMERAL ONE HUNDRED
xx14 RAQM NUMERAL TWO HUNDRED
xx15 RAQM NUMERAL THREE HUNDRED
xx16 RAQM NUMERAL FOUR HUNDRED
xx17 RAQM NUMERAL FIVE HUNDRED
XX18 RAQM NUMERAL SIX HUNDRED
XX19 RAQM NUMERAL SEVEN HUNDRED
xx1A RAQM NUMERAL EIGHT HUNDRED
xx1B RAQM NUMERAL NINE HUNDRED
xx1C RAQM NUMERAL ONE THOUSAND
xx1D RAQM NUMERAL TWO THOUSAND
xx1E RAQM NUMERAL THREE THOUSAND
xx1F RAQM NUMERAL FOUR THOUSAND
xx20 RAQM NUMERAL FIVE THOUSAND
xx21 RAQM NUMERAL SIX THOUSAND
xx22 RAQM NUMERAL SEVEN THOUSAND
xx23 RAQM NUMERAL EIGHT THOUSAND
xx24 RAQM NUMERAL NINE THOUSAND
xx25 RAQM NUMERAL TEN THOUSAND
xx26 RAQM NUMERAL TWENTY THOUSAND
xx27 RAQM NUMERAL THIRTY THOUSAND
xx28 RAQM NUMERAL FORTY THOUSAND
xx29 RAQM NUMERAL FIFTY THOUSAND
xx2A RAQM NUMERAL SIXTY THOUSAND
xx2B RAQM NUMERAL SEVENTY THOUSAND
xx2C RAQM NUMERAL EIGHTY THOUSAND
xx2D RAQM NUMERAL NINETY THOUSAND
xx2E RAQM NUMERAL ONE HUNDRED THOUSAND
xx2F RAQM NUMERAL TWO HUNDRED THOUSAND
xx31 RAQM NUMERAL TEN MILLION
xx32 RAQM NUMERAL TWENTY MILLION
xx33 RAQM NUMERAL COMBINING ONE
xx34 RAQM NUMERAL COMBINING TWO
xx35 RAQM NUMERAL COMBINING THREE
xx36 RAQM NUMERAL COMBINING FOUR
xx37 RAQM NUMERAL COMBINING FIVE
xx38 RAQM NUMERAL COMBINING SIX
xx39 RAQM NUMERAL COMBINING SEVEN
xx3A RAQM NUMERAL COMBINING EIGHT
xx3B RAQM NUMERAL COMBINING NINE
xx3C RAQM FRACTION ONE QUARTER
xx3D RAQM FRACTION ONE HALF
xx3E RAQM FRACTION THREE QUARTERS
xx3F RAQM CURRENCY MARK RUPEE
xx40 RAQM HUNDRED THOUSANDS MARK
xx41 RAQM PLACEHOLDER MARK
```


## 6 References

Ahmad Dihlavī, Sayyid. 1974. ;نْنَ آصغيت; [Farhang-i Āsafiyyah]. vol. 2. Delhi: Taraqqī-e-Urdū Board, National Academy.
Barker, Muhammad Abd-al-Rahman. 1967. A Course in Urdu. Vol. 1. Montreal: Institute of Islamic Studies, McGill University.
Gladwin, Francis. 1790. A Compendious System of Bengal Revenue Accounts. In three parts. Part I. Calcutta: Manuel Cantopher.
Naim, Chaudhry M. 1999. Introductory Urdu. 3rd ed. rev. Chicago: South Asia Language \& Area Center, University of Chicago.

Muhazzab, Muhammad Mirza. 195-?. مُّب اللغت [Muhazzab al-Lughat]. Lakhnau, Muhafiz Urdu Book Depot.
Muqtadirah Qaumi Zaban. 2001. ورّى اروو لغت [Darsi Urdu Lughat]. Taba‘ 1. Silsilah-yi matbu'at-yi Muqtadirah-yi Qaumi Zaban; 391. Islamabad.
Pandey, Anshuman. 2007a. "Proposal to Encode North Indic Number Forms in ISO/IEC 10646." ISO/IEC JTC1/SC2/WG2 N3367 L2/07-354. October 7, 2007. http://std.dkuug.dk/jtc1/sc2/wg2/docs/ n3367.pdf
—_ 2007b. "Proposal to Encode the Ganda Currency Mark for Bengali in ISO/IEC 10646." ISO/IEC JTC1/SC2/WG2 N3311R L2/07-192R. October 8, 2007. http://std.dkuug.dk/jtc1/sc2/wg2/ docs/n3311.pdf
Platts, John T. 1909. A Grammar of the Hindūstānī or Urdū Language. 5th imp. London: C. Lockwood. Stewart, Charles. 1825. Original Persian Letters and Other Documents with Fac-Similes. London: printed for the author by W. Nicol.


Figure 1: Table showing Siyaq forms as used in South Asia (from Platts 1909: 60). It should be noted that the translated values of the Raqm examples are wrong. The value 795 Rs and $11 \frac{3}{4} \mathrm{As}$ as represented in Raqm is actually 295 Rs and $1 \frac{3}{4}$ As. Platts represents $/$ as $1 / 2 \bar{a} n \bar{a}$ in the table, but assesses it as $3 / 4 \bar{a} n \bar{a}$ in the examples.

| SYMBOL | VALUE | SYMBOL | VALUE | SYMBOL | VALUE |
| :---: | :---: | :---: | :---: | :---: | :---: |
| , | $-/-/ 3$ | , - | -/-/9 | , 1 | -/1/3 |
| , - | $-/-/ 6$ | , 1 | -/1/- | 1 | -/1/6 |
| SYMBOL | VALUE | SYMBOL | VALUE | SYMBOL | VALUE |
| , 1 | -/1/9 | عـعـهِ | 12/-/- | , | 70/-/- |
| , | -/2/- | - | 13/-/- | , | 80/-/- |
| \% | 1/-/- | العسعه, | 14/-/- | لحسبر | 90/-/- |
| , C | 2/-/- | مـــــــ | 15/-/- | $\cdots$ | 100/- |
| , | 3/-/- |  | 16/-/- | $\Omega$ | 200/- |
| للعء, | 4/-/- |  | 17/-/- | $\omega$ | 300/- |
| عـ, | 5/-/- | , | 18/-/- | , لا | 400/- |
| $\cdots$ | 6/-/- | لعـعهو | 19/-/- | صما, | 500/- |
| , | 7/-/- | , | 20/-/- | , | 600/- |
| 1 | 8/-/- | , 0 | 30/-/- | معـا | 700/- |
| لهـ, | 9/-/- |  | 40/-/- | $\omega$ | 800/- |
| , | 10/-/- | مــه | 50/-/- |  | 900/- |
| , | 11/-/- | , 0 | 60/-/- | الـّه, | 1,000/- |
|  |  |  |  | $0$ | lakh/- |

Figure 2: Table showing Siyaq forms as used in South Asia (from Barker 1967: 356-357).
8.6. Sums: Both India and Pakistan now have a decimal coinage system, a rupee being divided into one hundred paisas. In Urdu, the decimal point is wirtten as: $s$ Examples:

$$
\text { is }=\text { Re. } 1.00 \quad s \Delta \cdot=50 \mathrm{p} . \quad s \cdot Q=5 \mathrm{p} . \quad \mid s \mathrm{I}^{\mathrm{r}}=\text { Rs. } 1.14
$$

8.7. Before the currency was refomed in the two countries, a rupee was divided into sixteen annas or sixty-four pice (paisa). There was then also a different system, besides the numerals, for writing sums.

$l_{6}=$ Rs. 2/-
$\omega=$ Rs. 3/-
, $=$ Rs. $4 /-\quad$ R Rs. $5 /-$

$$
, \angle=\text { Rs. } 6 /
$$

/ = Rs. 7/-

$$
, \infty=\text { Rs. 8/- }
$$

$$
\text { , }=\text { Rs. 9/- }
$$

$$
\text { , }=\text { Rs. 10/- }
$$

,
=Rs. 12/-

$$
\underline{\sim}=\text { Rs. 13/- }
$$

=لكعكــه = Rs. 14/-

$$
\text { ORs. } 1 \text { 5/- }
$$

, = Rs. 16/-
,
an=Rs. 18/-
K Rs. 19/-

$$
\text { , عe Rs. 20/- , , }=\text { Rs. } 30 /-
$$

, لا
=Rs. 50/-
, =Rs. 60/-
,
R = Rs. 80/-

$$
\text { لحـــهم Rs. } 90 /-
$$

$$
\Omega=\text { Rs. } 100 /-\quad \nearrow-=1 / 4 \text { anna or } 1 \text { pice }
$$

\[

\]

$$
\frac{\mu}{\mu}=\text { Rs. } 3 \text { and } 2 \text { annas } \& 3 \text { pice }
$$

Figure 3: Table showing Siyaq forms as used in South Asia (from Naim 1999: 49-50).
!


0

活
إِلْبُ





$$
\begin{aligned}
& 3 \\
& \text { عكُرil }
\end{aligned}
$$

Figure 4: The Arabic sources of the Urdu Siyaq forms (from Muhazzab 195-?: 51).


- (انگرينى) 1- 2- 3- 4- 5- 6- 7- 8- 9-10-
- 

ــــــ

$$
-x-i x-v i i i-v i i-v i-v-i v=i i i-i i-i(ر) \quad-
$$

- 

Figure 5: Table showing Siyaq forms as used in South Asia (from Muqtadirah Qaumi Zaban 2001: 718).


Figure 6: Table showing Siyaq forms as used in South Asia (from Dihlavi 1974: 363).

TABLE OF FIGURES. $\zeta$


Oblemve, that Annas are difinguifhed from Gundahs by the flroke being placed to the left of the former, and on the right fie of the latter.

Figure 7: Table showing method of writing fractions in South Asian tradition (from Gladwin 1790: 5)

Rو الامد


Figure 8: Table showing Siyaq forms as used in South Asia (from Stewart 1825: Plate 7).



| Rekern. | Hixd. |  | Rekem. | 景 |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | . | 500 | 1 | $\wedge$ | 81 |
| 6 | . | 600 | $\mathrm{O}_{1}$ | A | 82 |
| 4 | 6.0 | 700 | $\xrightarrow{\square}$ | iN | 83 |
| $1 /$ | ^.. | 800 | Q_J | A F | 84 |
| $r$ | $4 \cdot \cdot$ | 900 |  | is. | 85 |
| / | $1 \cdots$ | 100 | 0 | 14 | 86 |
| \| | r... | 2000 |  | ir | 7 |
|  | $\mu .$. | 2000 |  |  |  |
| \% لال | F... | 3000 |  | $\wedge \wedge$ | 88 |
| / | r'... | 4000 | Q | iq | 89 |
|  | -... |  | O_S | 9. | 0 |
| / |  |  |  |  |  |
|  | $4 .$. | 6000 | لر | 91 | 91 |
| \% | ᄂ... | 7000 | S | 9 F | 92 |
| / | A... | 8000 | 0 | $4 \mu$ | 93 |
| / | ¢,.. | 9000 | Q 3 |  | , |
| $\boldsymbol{\mu}$ |  |  |  |  | 94 |
| 18 | $1 \cdots$ | 10,000 | - | $9^{\circ}$ | 95 |
|  | F.... | 20,000 | Q | 94 | 96 |
| / | r.... | 30,000 | 0 | 94 | 97 |
| 10\% | r | 40,000 |  | Q | \% |
|  |  |  |  |  | 98 |
| 10 | 0.... | 50,000 | O-3 | 49 | 99 |
| 1 | y... | 60,000 |  |  |  |
| w |  |  |  |  | 100 |
| - | L. | 70,000 | $\Omega$ | F.. | 200 |
| / | A.... | 80,000 | 12 | $\mu \cdot$. | 300 |
| / | 9.... | 90,000 | $L$ | P'.. | 400 |

Figure 9: Table showing printed Siyaq forms as used in South Asia (from Gladwin 1790: 2-4).

The Rekem, or Siyak charallers, bcing only contraElions of Arabic words, the following Tible may firec to impriss them on the memory.



Figure 10: Table showing the Arabic sources of Siyaq forms (from Gladwin 1790: 6-7).


[^0]:    ${ }^{1}$ Gladwin 1790: vii.

