Proposal to Encode Oriya Fraction Signs in ISO/IEC 10646

Anshuman Pandey
University of Michigan
Ann Arbor, Michigan, U.S.A.
pandey@umich.edu

May 5, 2008

Contents

List of Figures

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: **Proposal to Encode Oriya Fraction Signs in ISO/IEC 10646**
2. Requester's name: **Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Individual contribution**
4. Submission date: **May 5, 2008**
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: **Oriya**
2. Number of characters in proposal: **6**
3. Proposed category: **A - Contemporary**
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 contains normalized forms of the Oriya fraction signs as found in printed documents. It was drawn by Anshuman Pandey using Metafont and converted to True Type format 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)? **No**
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, numeric information, and currency information 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 signs were used primarily in the state of Orissa, India.**
4. The context of use for the proposed characters (type of use; common or rare): **Common**
   (a) Reference: **The signs were used for numeric notation in the Oriya script.**
5. Are the proposed characters in current use by the user community?: **No.**
   (a) If Yes, where? Reference:
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?: **Yes**
   (a) If Yes, is a rationale provided?: **The signs belong to the Oriya script, which is encoded in the BMP. There is sufficient space in the Oriya block for the inclusion of the proposed signs.**
      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 constitute a specialized 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? **Yes**
    (a) If Yes, is a rationale for its inclusion provided? **Yes**
       i. If Yes, reference: **See text of proposal.**
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? **N/A**
       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**

1   Introduction

This is a proposal to encode six signs used for writing fractions in the Oriya script in the Basic Multilingual Plane (BMP) of the Universal Character Set (UCS) (ISO/IEC 10646).

The encoding of Oriya fraction signs is necessary for the full and accurate reproduction and representation of numeric notation in the Oriya script.

2   Characters Proposed

Characters   The six characters proposed are:

| ORIYA FRACTION ONE QUARTER

Ч   ORIYA FRACTION ONE HALF

Щ   ORIYA FRACTION THREE QUARTERS

╱   ORIYA FRACTION ONE SIXTEENTH

⁄   ORIYA FRACTION ONE EIGHTH

⁄   ORIYA FRACTION THREE SIXTEENTHS

Basis for Character Shapes   The forms of the proposed characters are based on normalized forms found in printed documents (see Figure 1 and Figure 2).

3   Technical Features

Name   The names of the fraction signs are based on the fraction values they represent.

Allocation   The characters should be encoded in the Oriya block (U+0B00).

Classification   The fraction signs may be categorized as a specialized set of characters belonging to a "Category A" (contemporary) script, as per the criteria specified in ISO/IEC JTC 1/SC 2/WG 2 N3002.[1]

Properties   The fraction signs belong to the Unicode general category "Number, Other" (No). Each character is assigned the numeric value of the fraction is represents. The signs are written left-to-right and are given bidirectional values of "Left-to-Right" (L).

```
0Bxx;ORIYA FRACTION ONE QUARTER;No;0;L;;;;1/4;N;;;;;
0Bxx;ORIYA FRACTION ONE HALF;No;0;L;;;;1/2;N;;;;;
0Bxx;ORIYA FRACTION THREE QUARTERS;No;0;L;;;;3/4;N;;;;;
0Bxx;ORIYA FRACTION ONE SIXTEENTH;No;0;L;;;;1/16;N;;;;;
0Bxx;ORIYA FRACTION ONE EIGHTH;No;0;L;;;;1/8;N;;;;;
0Bxx;ORIYA FRACTION THREE SIXTEENTHS;No;0;L;;;;3/16;N;;;;;
```

---

[1] International Organization for Standardization, 2005: 4.

4   Overview of the Signs

Description   The fraction signs were used for numeric notation in the Oriya script. The signs represent fraction values of a base-16 system. They appear in both written and printed materials. The characters were in common use until 1958, by which time the government of India had adopted the decimal system for currency and the metric system for weights and measures. The fraction signs are rarely used at present.

Typology   The six fraction signs can be divided into two sets: the quarter fractions ⎮ (¼), ᨐ(½), and ᨐ (¾); and the sixteenths ⟋ (¹⁄₁₆), ᨐ (⅛), and ᨐ (³⁄₁₆).

Orthography   Fractions are written using the additive principle. The signs for the quarter fractions are always written before the signs for the sixteenths fractions. For example, the fraction ⅜ is written as ⎮ᨐ, not as ᨐ⎮. All fraction values of the base-16 system can be expressed using the six proposed characters:

¹⁄₁₆	⟋	⁵⁄₁₆	⎮⟋	⁹⁄₁₆	ᨐ⟋	¹³⁄₁₆	ᨐ⟋
⅛	ᨐ	⅜	⎮ᨐ	⅝	ᨐᨐ	⅞	ᨐᨐ
³⁄₁₆	ᨐ	⁷⁄₁₆	⎮ᨐ	¹¹⁄₁₆	ᨐᨐ	¹⁵⁄₁₆	ᨐᨐ
¼	⎮	½	ᨐ	¾	ᨐ	1	ᨐ

5   Relationship to Other Signs

The Oriya fraction signs are analogous to those used in north Indic scripts and Bengali.[2] These regional orthographies all use six signs for representing values of a common base-16 system:

	⟋	ORIYA FRACTION ONE SIXTEENTH
¹⁄₁₆	⟋	U+09F4 BENGALI CURRENCY NUMERATOR ONE
	⎯	U+A833 NORTH INDIC FRACTION ONE SIXTEENTH
	ᨐ	ORIYA FRACTION ONE EIGHTH
⅛	ᨐ	U+09F5 BENGALI CURRENCY NUMERATOR TWO
	⹀	U+A834 NORTH INDIC FRACTION ONE EIGHTH
	ᨐ	ORIYA FRACTION THREE SIXTEENTHS
³⁄₁₆	ᨐ	U+09F6 BENGALI CURRENCY NUMERATOR THREE
	☰	U+A835 NORTH INDIC FRACTION THREE SIXTEENTHS
	⎮	ORIYA FRACTION ONE QUARTER
¼	⎮	U+09F7 BENGALI CURRENCY NUMERATOR FOUR
	⎮	U+A830 NORTH INDIC FRACTION ONE QUARTER
	ᨐ	ORIYA FRACTION ONE HALF
½	⎮⎮	U+09F7 BENGALI CURRENCY NUMERATOR FOUR (written twice)
	‖	U+A831 NORTH INDIC FRACTION ONE HALF
	ᨐ	ORIYA FRACTION THREE QUARTERS
¾	ᨐ	U+09F8 BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR
	⦀	U+A832 NORTH INDIC FRACTION THREE QUARTERS

---

[2] See Pandey (2007) for a description of north Indic fraction signs, to be encoded in the Common Indic Number Forms block.

Comparison of Signs   Of the north Indic and Bengali orthographies, the Oriya fraction signs are closest in appearance to Bengali forms. This is on account of the typological similarities between the Bengali and Oriya scripts, both being descendents of the Proto-Bengali family.

	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
ORIYA	╱	᭱	᭰	∣	Ч	Чᶗ
BENGALI	╱	᭰	᭰	∣	∣∣	᭰
NORTH INDIC	‾	=	≡	∣	∥	∥∣

As indicated above, three of the Oriya signs are graphically similar to their Bengali counterparts ($\frac{1}{16}$, $\frac{1}{8}$, and $\frac{1}{4}$), two are graphically distinct ($\frac{3}{16}$ and $\frac{3}{4}$), and one has no analogue in Bengali ($\frac{1}{2}$).[3] The signs for $\frac{1}{2}$ and $\frac{3}{4}$ are unique to Oriya.

- ╱ ORIYA FRACTION ONE SIXTEENTH is typologically and graphically similar to ╱ U+09F4 BENGALI CURRENCY NUMERATOR ONE and ‾ U+A833 NORTH INDIC FRACTION ONE SIXTEENTH. All three signs are stylistically distinct.

- ᭱ ORIYA FRACTION ONE EIGHTH is typologically and graphically similar to ᭰ U+09F5 BENGALI CURRENCY NUMERATOR TWO, but stylistically distinct. It is completely different from = U+A834 NORTH INDIC FRACTION ONE EIGHTH. The shapes of the Oriya and Bengali signs are derived by writing a stylized form of the respective digit 2 to the sign for $\frac{1}{16}$: Oriya ୨ → ᭱; Bengali ২ → ᭰.

- ᭰ ORIYA FRACTION THREE SIXTEENTHS is typologically similar, but graphically distinct from ᭰ U+09F6 BENGALI CURRENCY NUMERATOR THREE. This sign is typologically uniform among the scripts in the Proto-Bengali family; compare Oriya and Bengali to the Maithili form ᭰. In the eastern Indic scripts, the sign is derived by writing a stylized form of the digit 3 to the sign for $\frac{1}{16}$: Oriya ୩ → ᭰; Bengali ৩ → ᭰; Maithili ३ → ᭰.

- ∣ ORIYA FRACTION ONE QUARTER is typologically and graphically similar to ∣ U+09F7 BENGALI CURRENCY NUMERATOR FOUR and ∣ U+A830 NORTH INDIC FRACTION ONE QUARTER, but stylistically distinct.

- Ч ORIYA FRACTION ONE HALF is unique. In Bengali, the fraction $\frac{1}{2}$ is represented by writing ∣ U+09F7 BENGALI CURRENCY NUMERATOR FOUR twice; eg. the fraction $\frac{11}{16}$ is represented in Bengali as ∣∣᭰ and in Oriya as Ч᭰.

- Чᶗ ORIYA FRACTION THREE QUARTERS is distinct from Ч U+09F8 BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR and ∥∣ U+A832 NORTH INDIC FRACTION THREE QUARTERS.

---

[3] It is possible that the Bengali sign for the fraction $\frac{1}{2}$ might be an independent character consisting of two vertical bars, similar to ∥ U+A831 NORTH INDIC FRACTION ONE HALF. The author does not know if an independent sign for $\frac{1}{2}$ exists in Bengali or if the decision to not encode the sign independently was based on the principle of representing this character using primitives, ie. writing ∣ U+09F7 BENGALI CURRENCY NUMERATOR FOUR twice.

6 Justification for Encoding

The six proposed Oriya fraction signs constitute a specialized set of characters used for numeric notation and should be encoded as such in the UCS. Although the similarity of visual characteristics between certain Oriya and Bengali characters might suggest their candidacy for unification, other factors recommend against such action. These are differences in character names and property values.

Character Names　　The proposed Oriya characters are named according to their semantic value. The names of corresponding Bengali signs in the UCS possess the descriptor "currency numerator." For example the sign for ¼ in Oriya is represented by the character named ORIYA FRACTION ONE QUARTER, but by the character BENGALI CURRENCY NUMERATOR FOUR. The Bengali names are semantically unsound for characters that represent fractions and that are used for writing values of a range of units, not only currency. The descriptor "fraction" used in the names of the Oriya characters appropriately reflects the broader contexts of their use.

Numeric Value　　The numeric values assigned to the Oriya fraction signs represent their actual values. Those of the Bengali currency numerators are integers:

```
09F4;BENGALI CURRENCY NUMERATOR ONE;No;0;L;;;;1;N;;;;;
09F5;BENGALI CURRENCY NUMERATOR TWO;No;0;L;;;;2;N;;;;;
09F6;BENGALI CURRENCY NUMERATOR THREE;No;0;L;;;;3;N;;;;;
09F7;BENGALI CURRENCY NUMERATOR FOUR;No;0;L;;;;4;N;;;;;
09F8;BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR;No;0;L;;;;;N;;;;;
```

Integer numeral values are not appropriate for characters that fundamentally represent fractions. For example, ꠷ ORIYA FRACTION ONE EIGHTH and ꠶ U+09F5 BENGALI CURRENCY NUMERATOR TWO are used in the respective scripts to represent ⅛; however, the proposed Oriya character is assigned the numeric value ⅛, while the Bengali character possesses the numeric value 2.

If the differences between the three similar Oriya and Bengali are insufficient to encode the Oriya signs separately, it might be recommended that the numeric properties of the Bengali characters in question be redefined; for example, the numeric value of BENGALI CURRENCY NUMERATOR ONE should be changed from 1 to $\frac{1}{16}$, 2 to ⅛, etc. The drawback is that BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR would lose its implied generic value of n-1 (it presently lacks a numeric value) and would take the specific value of ¾. As a result, the name of the character would not match its numeric properties.

Historical Significance　　The six fraction signs are a historical component of the Oriya script and should be encoded to preserve the characters and to enable the digital encoding of records in which they appear.

7 References

International Organization for Standardization. 2005. "Principles and Procedures for Allocation of New Characters and Scripts." ISO/IEC JTC 1/SC 2/WG 2 N3002. October 5, 2005. `http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3002.pdf`.

Pandey, Anshuman. 2007. "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`

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

Young, A. H. 1935. *First Lessons in Oriya*. Revised by B. Das. Cuttack: Orissa Mission Press.

100   **NUMÉRATION OURIYA.**

CHIFFRES.	VALEURS.	NOMS DE NOMBRE.	CHIFFRES.	VALEURS.	NOMS DE NOMBRE.
୮୫	85	*pañtchâçi.*	୯୩	93	*téyânabé.*
୮୬	86	*tchhayâçi.*	୯୪	94	*tchourânabé.*
୮୭	87	*satâçi.*	୯୫	95	*pañtchânabé.*
୮୮	88	*aṭhâaçi.*	୯୬	96	*tchhayânabé.*
୮୯	89	*ounnabé.*	୯୭	97	*satânabé.*
୯୦	90	*nabé.*	୯୮	98	*aṭhânabé.*
୯୧	91	*ékânabé.*	୯୯	99	*ounçayé.*
୯୨	92	*byânabé.*	୧୦୦	100	*çayé.*

AUTRES NOMBRES PLUS ÉLEVÉS.

sahasr, hadjâr .. mille ;	*mahâkharb* .. 1 suivi de treize zéros ;
ayout .... dix mille ;	*çañkh* ...... 1 ..... quatorze zéros ;
lakch .... cent mille ;	*mahâçañkh* .. 1 ..... quinze zéros ;
niyouté ...... 1 suivi de six zéros ;	*hâhâ* .... 1 ..... seize zéros ;
koṭié .... 1 ...... sept zéros ;	*mahâhâhâ* ... 1 ..... dix-sept zéros ;
arboud .... 1 ...... huit zéros ;	*dhoûl* ...... 1 ..... dix-huit zéros ;
mahârboud .... 1 ...... neuf zéros ;	*mahâdhoûl* ... 1 ..... dix-neuf zéros ;
padma .... 1 ...... dix zéros ;	*akchôhiṇî* .... 1 ..... vingt zéros ;
mahâpadma ... 1 ...... onze zéros ;	*mahâkchôhiṇî.* 1 ..... vingt et un zéros.
kharb .... 1 ...... douze zéros ;	

FRACTIONS.

୲	$\frac{1}{4}$	*pâé* ou *pâddoulâ* ;	�१ ୳	$1\frac{1}{2}$	*dedh* ;
୳	$\frac{1}{2}$	*adh, ardh, arek* ;	୨ ୳	$2\frac{1}{2}$	*aḍhâi.*
୴	$\frac{3}{4}$	*paounou* ou *tinpâé* ;			

Figure 1: Oriya forms of the signs for the quarter fractions (from Pihan, 1860: 100).

Fractions.

The leading principle of Oriya arithmetic, to divide by *four* rather than any other number, pervades also the system of fractions.

ı means one quarter of the unit, rupee, or maund.

ꞁ „ two quarters.

ꟼ „ three quarters.

∕ „ one fourth of a quarter, or one sixteenth of the unit.

∕ „ two sixteenths.

∕ „ three sixteenths.

Figure 2: Oriya forms of the six fraction signs of the base-16 system (from Young, 1935: 203).