# 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**

	Proposal Summary Form	i
1	Introduction	1
2	Characters Proposed	1
3	Technical Features	1
4	Overview of the Signs	2
5	Relationship to Other Signs	2
6	Recommendation for Independent Encoding	4
7	References	4
Li	ist of Figures	
1		5
2	The six Oriya fraction signs depicted in an Oriya grammar book	6
3	Oriya notation for the base-16 fractions illustrated in an Oriya dictionary	6
4	Signs for Oriya quarter fractions shown in an Oriya grammar book	7

# ISO/IEC JTC 1/SC 2/WG 2 PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>

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.

<sup>&</sup>lt;sup>1</sup> 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 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 subset of Oriya.
- 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). This document (L2/08-199) is a revision of an earlier proposal (L2/07-413) and replaced the latter.

## 2 Characters Proposed

**Justification for Encoding** The six proposed Oriya fraction signs constitute a specialized set of characters in the Oriya script. The encoding of these fraction signs is necessary for the full and accurate reproduction and representation of numeric notation in Oriya.

**Characters** The six characters proposed are:

```
U+0B72 ORIYA FRACTION ONE QUARTER
U+0B73 ORIYA FRACTION ONE HALF
U+0B74 ORIYA FRACTION THREE QUARTERS
U+0B75 ORIYA FRACTION ONE SIXTEENTH
U+0B76 ORIYA FRACTION ONE EIGHTH
U+0B77 ORIYA FRACTION THREE SIXTEENTHS
```

**Basis for Character Shapes** The proposed characters are based on forms found in printed documents.

#### 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) at the range U+0B72..U+0B77.

**Classification** The fraction signs are a specialized subset of Oriya, a "Category A" (contemporary) script. <sup>1</sup>

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

```
0B72;ORIYA FRACTION ONE QUARTER;No;0;L;;;;1/4;N;;;;
0B73;ORIYA FRACTION ONE HALF;No;0;L;;;1/2;N;;;;
0B74;ORIYA FRACTION THREE QUARTERS;No;0;L;;;3/4;N;;;;
0B75;ORIYA FRACTION ONE SIXTEENTH;No;0;L;;;1/16;N;;;;
0B76;ORIYA FRACTION ONE EIGHTH;No;0;L;;;1/8;N;;;;
0B77;ORIYA FRACTION THREE SIXTEENTHS;No;0;L;;;3/16;N;;;;
```

**Line-Breaking Properties** A sequence of fraction signs must not be broken at line boundaries.

<sup>&</sup>lt;sup>1</sup> As per the criteria specified in ISO/IEC JTC 1/SC 2/WG 2 N3002 (International Organization for Standardization, 2005: 4).

## 4 Overview of the Signs

**Description** The characters were used for fraction notation in the Oriya script. They appear in both written and printed materials, and were cut as part of at least four different Oriya metal fonts, as represented in the figures. The fraction signs were in common use until 1958, when the Government of India adopted the decimal system for currency and the metric system for weights and measures. The characters are rarely used at present.

**Typology** The fraction signs can be divided into two sets: signs for the quarter increments —  $|(\frac{1}{4}), \frac{1}{4}(\frac{1}{2}), \frac{1}{4}(\frac{1}{4})|$  and signs for the sixteenths increments —  $|(\frac{1}{4}), \frac{1}{4}(\frac{1}{4})|$ .

**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  $\frac{3}{8}$  is written as  $\frac{1}{9}$ , not as  $\frac{9}{1}$ . Fraction values of the base-16 system are written using the six proposed characters as:

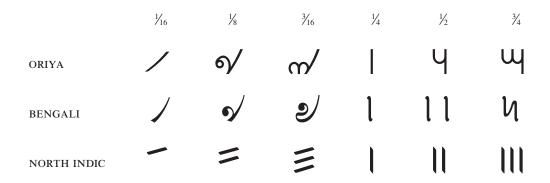
## 5 Relationship to Other Signs

The Oriya fraction signs are analogous to those used in north Indic scripts and Bengali:<sup>2</sup>

- / U+0B75 ORIYA FRACTION ONE SIXTEENTH
- $\frac{1}{16}$  J U+09F4 BENGALI CURRENCY NUMERATOR ONE
  - U+A833 NORTH INDIC FRACTION ONE SIXTEENTH
  - ∘/ U+0B76 ORIYA FRACTION ONE EIGHTH
- 1/8 v u+09F5 bengali currency numerator two
  - = U+A834 NORTH INDIC FRACTION ONE EIGHTH
  - ന∕ U+0B77 ORIYA FRACTION THREE SIXTEENTHS
- $\frac{3}{16}$  2 U+09F6 BENGALI CURRENCY NUMERATOR THREE
  - **■** U+A835 NORTH INDIC FRACTION THREE SIXTEENTHS
  - U+0B72 ORIYA FRACTION ONE QUARTER
- 1/4 U+09F7 BENGALI CURRENCY NUMERATOR FOUR
  - U+A830 NORTH INDIC FRACTION ONE QUARTER
  - 4 U+0B73 ORIYA FRACTION ONE HALF
- ½ 11 U+09F7 BENGALI CURRENCY NUMERATOR FOUR (written twice)
  - II U+A831 NORTH INDIC FRACTION ONE HALF
  - Ч U+0B74 ORIYA FRACTION THREE QUARTERS
- $^{3}\!\!\!/$  U +09F8 BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR
  - III U+A832 NORTH INDIC FRACTION THREE QUARTERS

<sup>&</sup>lt;sup>2</sup> See Pandey (2007) for details on north Indic fraction signs, to be encoded in the Common Indic Number Forms block (U+A830).

**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.



As indicated above, three of the Oriya signs are graphically similar to their Bengali counterparts ( $\frac{1}{6}$ ,  $\frac{1}{6}$ , and  $\frac{1}{4}$ ), two are graphically distinct ( $\frac{3}{6}$  and  $\frac{3}{4}$ ), and one has no analogue in Bengali ( $\frac{1}{2}$ ). The signs for  $\frac{1}{2}$  and  $\frac{3}{4}$  are unique to Oriya.

- / U+0B75 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.
- $\circ$ / U+0B76 ORIYA FRACTION ONE EIGHTH is typologically and graphically similar to  $\circ$ / U+09F5 BENGALI CURRENCY NUMERATOR TWO, but stylistically distinct. It is completely different from  $\leq$  U+A834 NORTH INDIC FRACTION ONE EIGHTH. The shapes of the Oriya and Bengali signs are derived by writing a stylized and rotated form of the respective digit 2 to the sign for  $\frac{1}{16}$ : Oriya  $9 + \frac{1}{16} \rightarrow 0$ ; Bengali  $\approx 1$
- $\omega'$  U+0B77 ORIYA FRACTION THREE SIXTEENTHS is typologically similar, but graphically distinct from  $\vartheta$  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  $\vartheta$ . In the eastern Indic scripts, the sign is derived by writing a stylized and rotated form of the digit 3 to the sign for  $\frac{1}{16}$ : Oriya  $\mathfrak{I} + \mathcal{I} \to \omega'$ ; Bengali  $\mathfrak{I} + \mathcal{I} \to \mathfrak{I}$ .
- | U+0B72 ORIYA FRACTION ONE QUARTER is typologically and graphically similar to | U+09F7 BEN-GALI CURRENCY NUMERATOR FOUR and | U+A830 NORTH INDIC FRACTION ONE QUARTER, but stylistically distinct.
- 4 U+0B73 ORIYA FRACTION ONE HALF is unique. In Bengali, the fraction ½ is represented by writing 1 U+09F7 BENGALI CURRENCY NUMERATOR FOUR twice; eg. the fraction ½ is represented in Bengali as 11 2/2 and in Oriya as 4 m/.
- <sup>U</sup>H U+0B74 ORIYA FRACTION THREE QUARTERS IS distinct from <sup>I</sup>M U+09F8 BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR and III U+A832 NORTH INDIC FRACTION THREE QUARTERS.

<sup>&</sup>lt;sup>3</sup> It is possible that the Bengali sign for the fraction ½ 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 ½ 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 Recommendation for Independent Encoding

The graphical similarity between certain Oriya and Bengali signs might suggest unification of Oriya signs with those of the latter. However, incompatibilities in character names and property values recommend against such action. The Oriya fraction signs should be encoded independently in the UCS.

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, of U+0B73 ORIYA FRACTION ONE EIGHTH and of U+09F5 BENGALI CURRENCY NUMERATOR TWO are used in the respective scripts to represent ½; however, the proposed Oriya character is assigned the numeric value ½, 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 currency numerators be redefined; for example, the numeric value of BENGALI CURRENCY NUMERATOR ONE should be changed from 1 to  $\frac{1}{16}$ , 2 to  $\frac{1}{8}$ , 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  $\frac{3}{4}$ . As a result, the name of the character would not match its numeric properties.

#### 7 References

Grundy, R. J. 1928. The Concise Oriya-English Dictionary. Cuttack: Orissa Mission Press.

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.

Sutton, Amos. 1831. *An Introductory Grammar of the Oriya Language*. Cuttack: Baptist Mission Press. 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.
<b>に</b> %	85	pañtcháç <b>í</b> .	4m	93	téyânabé.
∟ക	86	tchhayáçî.	የ	94	tchouránabé.
「つ	87	satáçi.	4%	95	pañtchânabé.
$\Gamma\Gamma$	88	aịhâaçî.	৮৩	96	tchhayánabé.
<u>г</u> е	89	ounnabé.	৫৯	97	satánabé.
५०	90	nabé.	4	98	a!hânabé.
48	91	ékánabé.	44	99	ounçayé.
48	99	byánabé.	१००	100	çayê.

#### AUTRES NOMBRES PLUS ÉLEVÉS.

sahasr, hadjár	mille;	mahdkharb	1 suivi de treize séros;
ayout	dix mille;	çañkh	1 quatorze zéros;
lakch	cent mille;	mahâçañkh	1 quinze zéros;
niyouté	1 suivi de six zéros;	hdhd	1 seize zéros;
koțié	1 sept zéros;	maháháhá	1 dix-sept zéros;
arboud	1 huit zéros;	dhoùl	1 dix-huit zéros;
mahårboud	1 neuf zéros;	mahâdhoùl	1 dix-neuf zéros;
padma	1 dix zéros;	akchóhint	1 vingt zéros;
mahápadma	1 onze zéros;	mahákchóhint.	1 vingt et un z4-
kharb	1 douze zéros;		ros.

#### FRACTIONS.

Figure 1: Signs for the Oriya quarter fractions shown in *Exposé des signes de numération*, a reference manual for number systems (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.

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

- 1 ,, two quarters.
- 4 ,, three quarters.
- / ,, one fourth of a quarter, or one sixteenth of the unit.
- , two sixteenths.
- ,, three sixteenths.

Figure 2: The six Oriya fraction signs depicted in First Lessons in Oriya (from Young, 1935: 203).

## Fractions.

GIA or GI, a quarter, is thus r	ed	•••	1	
ସଧ, ଅଧେ, ଅର୍ବ, ଅର୍ବେହ, half	•••	•••	•••	9
ସର୍ଜ୍ୱ, or ଜନସା, three-quarter	·s	•••	•••	લ
coo, one and a half	•••	•••	•••	५१
TOIQ, two and a half	•••	•••	•••	99

These are irregular. In all other cases a quarter more than even number is expressed by prefixing qqq, to the number; a half by prefixing qqq; and three-quarters by prefixing qqq to the even number. Exam. qqq qq, three and a quarter; qqq qq, three and a half; qqqq qq, three and three-quarters, viz., a quarter less than four.

Fractional parts are generally expressed by dividing the Rupee into sixteen annas or parts. *Exam.* **QIQ QAI**, seven annas, or seven-sixteenths; **QA QAI**, ten annas, or ten-sixteenths.

## The signs are as follow:-

1	•∕	<b>~</b>	ı	1	l•∕	l ee/	٩	9/	40/	9~	4	4/	4€	44	•
16	8	3	<u>t</u>	<u>5</u>	6 16	7_ 16	<u>I</u>	<u>9</u>	16	11	3	13	14	15	1

Figure 3: Specimen from *The Concise Oriya-English Dictionary* illustrating the Oriya notation for the base-16 fractions (from Grundy, 1928: 11). The text is the same as the description given for fractions in Sutton's 1831 grammar (see Figure 4), but the font used is different.

# (56)

#### FRACTIONS.

ঘাৰ্থ or ঘাতুন্তা, is a quarter, and thus represented,
খ্য, খ্য, খ্রেক, is half,
ঘন্ত্র, or ভিন্নঘাৰ, is three quarters,
দেও, is one and a half,
খ্যান্ত, two and a half,

These are irregular. In all other cases a quarter more than an even number is expressed, by prefixing sequences; a half, by prefixing sign and three quarters, by prefixing sequences, to the next even number. Exam. sequences and a quarter, signs three and a quarter, signs three and a half, segs three and three quarters, viz. a quarter less than four.

Fractional parts are generally expressed by dividing the Rupee into sixteen Anas, or parts. Exam. হাভেখনা, seven Anas, or seven-sixteenths; হ্যাখনা ten Anas, or ten-sixteenths.

Figure 4: Specimen from *An Introductory Grammar of the Oriya Language* showing signs for the Oriya quarter fractions (from Sutton, 1831: 56).