

- UNICODE is emerging as international de-facto industry standard. It is 2-byte code. The Unicode consortium provides a number of on-line resources for obtaining information and data about the Unicode standard, as well as updates and corrigenda. These are:
Unicode Web site: <http://www.unicode.org>
Unicode Public Mailing List: unicode@unicode.org
- The following are of our relevance in the UNICODE standard version 3.0.
 - (a) IPA extensions (0250-02AF) on PP 359-367
 - (b) Devanagari (0900 - 097F), Bengali (0980 - 09FF), Gurmukhi (0A00 - 0A7F), Gujarati (0A80 - 0AFF), Oriya (0B00 - 0B7F), Tamil (0B80 - 0BFF), Telugu (0C00 - 0C7F), Kannada (0C80 - 0CFF), Malayalam (0D00 - 0D7F), on pp 211-235 and 401-425
- Observations and comments :
 - (i) Sinhala (0D80 - 0DFF), Tibetan (0F00 - 0FBF), and Myanmar (1000 - 109F) follow the alphabetic order similar as in DEVANAGARI and other Indian scripts.
 - (ii) Elaborate description for DEVANAGARI (U+0900 - U+097F) is given on pp 211 - 223. This needs review in view of some additions/revisions. This includes sections on Standards, Encoding Principles, Rendering mechanism, Consonant letters, Independent Vowel letters, Dependent Vowel signs (matras), Viram, consonants conjuncts, Rules for Rendering (R1 - R14), Memory Representation and Rendering order, sample half-forms and ligatures, combining marks, digits and punctuation symbols.
 - (iii) Use of VIRAM in Unicode 3.0 is wrong. This must be replaced by HALANT. VIRAM means full stop as punctuation mark. Halant suppresses inherent vowel, and thus used with consonant to represent pure consonant.
 - (iv) Rendering mechanism is suggested as follows :

(a) Linear conjunction

क	्	व	= क्व
0915	094D	0935	
क	्	ष	= क्ष
0915	094D	0937	
त	्	र	= त्र
0924	094D	0930	
ज	्	ञ	= ज्ञ
091C	094D	091E	

(b) Vertical conjunction

क	व्	व	= क्व
0915	200D	0935	

(v) Last column of the code-set of an Indian script is proposed to be used for assigning language-specific letters and signs such as currency sign, pure vowel, consonant signs etc.

(vi) As for Devanagari similar detailed write-up is to be prepared for each of other Indian Scripts also such as Bengali (covering Assamese and Manipuri as well), Gurmukhi, Gujarati, Oriya, Telugu, Kannada, Malayalam. Description of Tamil needs to be revisited. Detailed description may include the section headings as follows:

Standards, Encoding Principles, Consonants, Vowels, Vowel modifiers, Conjuncts, Rules of Rendering, Ligatures, Combining marks, Digits and Punctuation symbols.

(vii) Both the National standard and the Unicode standard will co-exist. It is expected that the coordinators of the Resource Centers for Indian Language Technology Solutions (RC-ILTS) will convene discussion meetings with representative(s) of the State Government and the language expert, and finalize the updates for UNICODE standard (version 3.0) on priority basis.

(viii) Transliteration among Indian languages should be simple. One- to- one code-point correspondence is desirable. Southern languages often use pure consonants. Gurmukhi uses pure vowels as well. This issue of representing pure consonants and pure vowels is not being raised at this stage for inputs to UNICODE.

(ix) Updating constraints of Unicode consortium regarding character encoding stability:

- Once a character is encoded, it will not be moved or removed.
- Once a character is encoded, its character name will not be changed.
- Once a character is encoded, its canonical combining class and decomposition (either canonical or compatibility) will not be changed in a way that would affect normalization.
- Once a character is encoded, its properties may still be changed, but not in such a way as to change the fundamental identity of the character.
- The structure of certain property values in the Unicode Character Database will not be changed.

● Preliminary feedback comments from the linguists and representatives of state governments are as follows :

■ DEVANAGARI (covering languages: Hindi, Marathi, Nepali, Sanskrit, Konkani, Sindhi). Unicode code-set (0900-097F)

DEVANAGARI code set on Page 401 with reference to this, the following updates are proposed.

	090	091	092	093	094	095	096	097
0		ऐ 0910	ठ 0920	र 0930	ी 0940	ॐ 0950	ॠ 0960	० 0970
1	ँ 0901	ऑ 0911	ड 0921	ॠ 0931	ु 0941	ं 0951	ॡ 0961	ॠ 0971
2	ं 0902	ओ 0912	ढ 0922	ल 0932	ू 0942	॑ 0952	ॢ 0962	
3	ः 0903	ओ 0913	ण 0923	ळ 0933	ृ 0943	े 0953	ॣ 0963	
4	ऐ 0904	औ 0914	त 0924	ळ 0934	॥ 0944	ँ 0954	। 0964	च 0974
5	अ 0905	क 0915	थ 0925	व 0935	ँ 0945	ॡ 0955	॥ 0965	
6	आ 0906	ख 0916	द 0926	श 0936	ँ 0946	ँ 0956	० 0966	
7	इ 0907	ग 0917	ध 0927	ष 0937	ँ 0947	ॡ 0957	१ 0967	
8	ई 0908	घ 0918	न 0928	स 0938	ै 0948		२ 0968	
9	उ 0909	ङ 0919	न 0929	ह 0939	ँ 0949		३ 0969	
A	ऊ 090A	च 091A	प 092A	० 093A	ौ 094A		४ 096A	
B	ॠ 090B	छ 091B	फ 092B		ौ 094B		५ 096B	
C	ॡ 090C	ज 091C	ब 092C	ॠ 093C	ौ 094C		६ 096C	
D	ँ 090D	झ 091D	भ 092D	ॠ 093D	ॠ 094D		७ 096D	
E	ऐ 090E	ज 091E	म 092E	ा 093E			८ 096E	
F	ए 090F	ट 091F	य 092F	ि 093F			९ 096F	

Additions (in the available code-space)

- 0904 अे DEVANAGARI LETTER SHORT A
 093A ० DEVANAGARI INVISIBLE LETTER
 0955 ॠ DEVANAGARI ANUSWARA
 . Used in Sanskrit Yajurveda
 0956 ॡ DEVANAGARI SHUKLA YAJURVEDIC ANUSWARA
 . Used in Sanskrit
 0957 ॢ DEVANAGARI JIVHYAMULIYA
 . Used in Sanskrit
 0958 to 095 F Usage of these code-points be discouraged,
 and rather use the corresponding
 DEVANAGARI LETTER followed by
 Nukta sign 093C to generate sounds in
 Urdu, Sindhi etc. For example
 0958 क ≡ 0915 क 093c ्

Change in naming character

- 094D ॣ DEVANAGARI SIGN HALANT
 . Suppresses inherent vowel
 . with consonant to represent pure consonant

Change in explanation :

- 0970 ० DEVANAGARI ABBREVIATION
 SIGNS as in डॉ०

Language Specific Comments

(Note : RC-ILTS : Resource Centre for Indian Language
 Technology Solutions)

● HINDI

Representative Experts:

Dr. Pushplata Taneja , Central Hindi Directorate,
 MHRD, New Delhi. Tel: 011-6105211-14

Dr. Vijay Goel , Dept. of Official Languages,
 MHRD, New Delhi.
 Tel: 011-4619860 E-mail: vijay54@hotmail.com

Prof. Sanjay Dhande, Dean (R&D)
 Coordinator RC-ILTS, IIT, Kanpur
 Tel: 0512-598570 E-mail: sgd@iitk.ac.in

● MARATHI

Representative Experts:

Dr. S.P.Mudur, NCST, Mumbai.
 Tel: 022-6201606 E-mail: mudur@ncst.ernet.in

Prof. Pushpak Bhattacharya, Coordinator
 RC-ILTS, IIT, Mumbai
 Tel: 022-5767718 E-mail: pb@cse.iitb.ernet.in

● KONKANI

There are additional sounds such as

च, ज, टॉ which are realisable with use of Nukta (्) or
 Chandra (ँ)

च = च [091A] ँ [093C]

टॉ = ट [091F] ऌ [0949]

Representative Experts:

Prof. K.J.Mahale, Asmitai Pratisthan, Goa
 Tel:0832-222781 E-mail: mahale@goal.dot.net.in

Dr. Chandrelekha, Konkani Department, Goa
 University. Tel: 0832-514992

Prof. Pushpak Bhattacharya, Coordinator
 RC-ILTS, IIT, Mumbai. Tel: 022-5767718
 E-mail: pb@cse.iitb.ernet.in

● NEPALI

Representative Experts:

Shri K. Jaykumar, Dy. Secy, Dept. of AR&PG,
 Sardar Patel Bhawan, New Delhi Tel:011-3733030
 E-mail: kjayakumar@mailcity.com

Prof. Sanjay Dhande, Dean (R&D), Coordinator
 RC-ILTS, IIT, Kanpur Tel: 0512-598570
 E-mail: sgd@iitk.ac.in

● SINDHI

Additions :

1. 0974 च = DEVANAGARI LETTER SHORT YA
 . used in SINDHI

Additional Comments:

1. There is use of implosive consonants

ज यु जु ञ
 ञ डू ढू ञू

These are realisable using existing code-points; e.g.

जु = ज [091C] ू [0952] ु [0941]
 ञू = ज [091C] ् [093C] ू [0942]

2. 0952 ॡ = DEVANAGARI STRESS SIGN ANUDATTA
 . used in Sindhi to generate implosive
 consonants especially with

ग [0917], ज [091C], ड [0921], द [0926], ब [092C]]

Representative Experts:

Dr. Kishore Vaswani,
 National Council for Promotion of Sindhi Language , Vadodara
 Tel: 0265-342246
 Shri M.K.Jetley, Consultant Linguist, Preet Vihar, New Delhi
 Tel:011-2146121

Shri S. S. Pujari, Coordinator, RC-ILTS, C-DAC, Pune
 Tel: 020-5655228 E-mail: pujari@cdac.ernet.in

■ **BENGALI** Unicode code-set (0980 - 09FF)

This code-set is also used for Assamese and Manipuri languages.

Additions:

09BD:	১	: BENGALI SIGN AVAGRAH
09E4:	।	: BENGALI SIGN VIRAM DANDA
09E5:	॥	: BENGALI SIGN DOUBLE DANDA
09BA:	০	: BENGALI INVISIBLE LETTER
09CE:	০৫	: BENGALI YA
09CF:	০৬	: BENGALI RA
09DE:	০৭	: BENGALI TA

Change in Annotation:

09F4:		BENGALI CURRENCY NUMERATOR ONE . not in current usage
09F5:		BENGALI CURRENCY NUMERATOR TWO . not in current usage
09F6:		BENGALI CURRENCY NUMERATOR THREE . not in current usage

Representative Experts:

Prof. B.B.Choudhary, ISI, Kolkata
Tel: 033-5778085 E-mail: bbc@isical.ernet.in

Shri Tamal Sen, Webel, Kolkata
Tel: 033-3573700 E-mail: webel@giascl01.vsnl.net.in

■ **GURMUKHI** Unicode code-set (0A00 - 0A7F)

Additions :

0A01:	ੴ	: GURMUKHI SIGN ADAK BINDI . Nasalization
0A03:	ੳ	: GURMUKHI VISARG
0A3A:	ੴ	: GURMUKHI INVISIBLE LETTER
0A50:	ੴ	: GURMUKHI EK ONKAR . God is one
0A4E:	ੴ	: GURMUKHI PARI - RA
0A4F:	ੴ	: GURMUKHI PARI - HA
0A64:	।	: GURMUKHI SIGN DANDA or VIRAM
0A65:	॥	: GURMUKHI SIGN DOUBLE DANDA
0A78:	ਖ਼	: GURMUKHI SIGN KHANDA

Change in shape :

0A33: from ਝ to ਝ

Change in Interpretation:

0A74 : not in use; moved to 0A50

Representative Experts:

Shri Prabhjeet Singh, IS&AR, Govt. of Punjab, Chandigarh
Tel: 0172-604892

Dr. G.S.Lehal, Coordinator, RC-ILTS, Thapar
Institute for Engineering & Technology, Patiala
Tel: 0175-214498 E-mail: gslehal@mailcity.com

■ **GUJARATI** Unicode code-set (0A80 - 0AFF)

Additions :

0A8C :	૨	: GUJARATI LETTER VOCALIC L
0AD1:	ૠ	: GUJARATI STRESS SIGN UDAATTA . Used with Sanskrit text
0AD2 :	ૡ	: GUJARATI STRESS SIGN ANUDAATTA . Used with Sanskrit text
0AD3 :	ૢ	: GUJARATI GRAVE ACCENT SWARIT . Used with Sanskrit text
0AD4 :	ૣ	: GUJARATI ACUTE ACCENT SWARIT . Used with Sanskrit text
0ABA :	૦	: GUJARATI INVISIBLE LETTER
0AE1 :	૨	: GUJARATI LETTER VOCALIC LL . Used with Sanskrit text
0AE2 :	૩	: GUJARATI VOWEL SIGN VOCALIC L . Used with Sanskrit text
0AE3 :	૪	: GUJARATI VOWEL SIGN VOCALIC LL . Used with Sanskrit text
0AE4 :	।	: GUJARATI VIRAM or DANDA
0AE5:	॥	: GUJARATI DOUBLE DANDA
0AF1 :	₹	: GUJARATI RUPEE CURRENCY SIGN

Representative Experts:

Shri D.K.Sarvaiya, IT Dept., Govt. of Gujarat, Gandhinagar
Tel: 02712-50464 E-mail: dksarvaiya@hotmail.com

Dr. M.N.Cooper, Modular Infotech Ltd., Pune
Tel: 020-4227994 E-mail: modular@gspn01.vsnl.net.in

Shri Shitansu Y Mehta, Coordinator
RC-ILTS,MS Univ., Baroda
Tel: 0265-792959 E-mail: rciltg@satyam.net.in

■ **ORIYA** UNICODE code-set : (0B00 - 0B7F)

Comments

1. Shapes of following code-points must be changed:

0B0A :	ୠ	: ORIYA LETTER UU
0B0B:	ୡ	: ORIYA LETTER VOCALIC R
0B48:	ୢ	: ORIYA VOWEL SIGN AI
0B4C:	ୣ	: ORIYA VOWEL SIGN AU
0B66:	୦	: ORIYA DIGIT ZERO

Size of zero (0B66) be smaller than the size of letter TTHA (0B20)

ORIYA ANUSWAR: ୦ TTHA: ୦ ZERO: ୦

2. Change in Explanation

Annotation under 0B2C, i.e. in "Oriya va, wa", wa must be deleted.

3. Addition of characters

0B35: ୨ : ORIYA LETTER VA
0B64: ୧ : ORIYA VIRAM/DANDA
0B65: ୧୧ : ORIYA DOUBLE DANDA
0B3A: ୦ : ORIYA INVISIBLE LETTER

Representative Experts:

Dr. Sanghmitra Mohanty, Utkal Univ., Bhubaneswar
Tel. : 0674-580216 E-mail : sanghaw@dte.vsnl.net.in

Dr. A.K Pujari CEO, OCAC, Bhuwaneshwar
Tel : 0674-540830 E-Mail : akp@ocac.ernet.in

■ TAMIL UNICODE Code-set: 0B80 - 0BFF

Change in shape

1. 0B83 : This is "Ayatham" and it is represented as ୦.
It is an independent symbol.

Change in annotation

1. 0B82 : ୦ : TAMIL SIGN ANUSVARA
It is not used in Tamil
2. Explanations given under OBCA, OBCB and OBCC reflect equivalence of the vowel- modifiers as concatenated string of split modifiers.

The following explanations are wrong, as vowel modifier splitting equivalence is not correct. Hence the following explanations must be deleted.

OBBC ୦୧୩ ≡ OBC ୦୧
OBBC ୦୧୩ ≡ OBC ୦୧
OBBC ୦୧୩ ≡ OBC ୦୧

3. However, TAMIL letter sequencing as in the Unicode Standard 3.0 is also not acceptable. New code-set is being worked out.

Representative Experts:

Prof. M. Ponnaivaiko, Tamil Virtual University, Chennai. Tel: 044-2541010
E-mail: tamilvu@md3.vsnl.net.in

Dr. T.V.Geetha & Dr. Ranjani Parthasarathi,
Anna univ., Chennai
Tel:044-2351723 E-mail: rp@annauniv.ernet.in

■ TELUGU Unicode code-set (0C00 - 0C7F)

Additions:

0C3A: ୦ : TELUGU INVISIBLE LETTER
0C3D: ୫ : TELUGU SIGN AVAGRAH
0C3C: ୦୧ : TELUGU SIGN NUKTA
. placed at bottom left corner of the letter
0C0D : ୦୦ : TELGU SIGN AYE
0C11 : ୧୦ : TELGU SIGN AWE
0C34: ୧୧ : TELGU SIGN ZHA

Change in Naming

0C4D: ୧ : TELUGU SIGN HALANT
. This is Halant to take out hidden vowel sound of the consonant.

Representative Experts:

Shri J. Satyanarayana,
Secretary to Govt. of Andhrapradesh, Hyderabad
Tel: 040-3456401 E-mail : secy_it&c@ap.gov.in

Prof. K. Narayanmurthy, Coordinator,
RC-ILTS, Univ. of Hyderabad, Hyderabad
Tel: 040-3010500 E-mail: knmcs@uohyd.ernet.in

■ KANNADA Unicode code-set (0C80 -0CFF)

Additions:

0CBA : ୦ : KANNADA INVISIBLE LETTER
0CBC : ୦୧ : KANNADA SIGN NUKTA
0CBD : ୫ : KANNADA AVAGRAHA
0CF8 : ୧ : KANNADA GLYPH SIGN
0CD2 : ୦୧ : KANNADA DIACRITIC SIGN ANUDATTHA
0CD1 : ୦ : KANNADA DIACRITIC SIGN SWARITHA
0CF9 : ୧ : KANNADA DIACRITIC SIGN DEERGA SWARITHA
0CD3 : ୦ : KANNADA DIACRITIC SIGN GURU
0CD4 : ୧ : KANNADA DIACRITIC SIGN LAGHU

Change in Annotation

Delete equivalence annotations, as splitting of vowel modifiers is not permitted.

under OCC0 [= OCC6 OCD5] to delete
under OCC7 [= OCC6 OCD5] to delete
under OCC8 [= OCC6 OCD6] to delete
under OCCA [= OCC6 OCC2] to delete
under OCCB [= OCCA OCD5] to delete

Representative Experts:

Shri C.V Srinath Sastry, Govt. of Karnatka, Bangalore
Tel : 080-2266291-93 E-mail: sastry@ead.cmmacs.ernet.in
Prof. N.J. Rao, Coordinator, RC-ILTS, IISc, Bangalore
Tel: 080-3092377 E-mail: njr@mgmt.iisc.ernet.in

■ MALAYALAM Unicode code-set (0D00 - 0D7F)

Additions:

0D78:	൩	: Malayalam letter NN
0D79:	൪	: Malayalam letter N
0D7A:	൫	: Malayalam letter RR
0D7B:	൬	: Malayalam letter L
0D7C:	൭	: Malayalam letter LL
0D73:	൮	: Malayalam Consonant sign YA
0D74:	൹	: Malayalam Consonant sign RRA
0D75:	ൺ	: Malayalam Consonant sign LLA
0D76:	ൻ	: Malayalam Consonant sign VA
0D3A:	ർ	: MALAYALAM INVISIBLE LETTER

Shape Change:

0D4C:	ൽ	: MALAYALAM VOWEL AU
-------	---	----------------------

Change in Explanation:

0D1A:	ച	: Malayalam letter CHA
0D1B:	ചഃ	: Malayalam letter CHHA
0D1E:	ഞ	: Malayalam letter NHA
0D1F:	ട	: Malayalam letter TA
0D20:	ത	: Malayalam letter TTA
0D21:	തഃ	: Malayalam letter (hard) DA
0D22:	തഃ	: Malayalam letter (hard) DDA
0D23:	ണ	: Malayalam letter (hard) NA
0D24:	ത	: Malayalam letter THA
0D25:	ഥ	: Malayalam letter TTHA
0D26:	ദ	: Malayalam letter (soft) DA
0D27:	ധ	: Malayalam letter (soft) DDA
0D34:	ഴ	: Malayalam letter ZHA
0D36:	ശ	: Malayalam letter (soft) SHA
0D37:	ഷ	: Malayalam letter SHA
0D4D:	ഃ	: Malayalam sign Chandrakkala
0D57:	ൿ	: Malayalam vowel sign AU

(Already given at 0D4C, should be deleted)

Representative Experts:

Ms. Ishita Roy, Additional Secy. IT Department,
Govt. of Kerala, Trivendrum
Shri Sassi P.M. Coordinator Mission on IT, Trivendrum
Tel: 0471-324277 E-mail: itmission@vsnl.com

Dr. Ravindra Kumar, Co-ordinator
RC-ILTS, ER&DC, Trivendrum
Tel: 0471-320116 E-mail: ravi@erdcitvm.org

■ ARABIC Unicode code-set (0600 - 067F)

In the context of Indic Scripts cover Urdu, Sindhi and Kashmere

Additions:

0656:	ﻻ	: ARABIC LETTER ULTA PESH . used in Urdu
0657:	ﻻ	: ARABIC LETTER JAZM .used in Urdu
0658:	ﻻ	: ARABIC LETTER KHARI ZER . used in Urdu
0659:	ﻻ	: ARABIC LETTER HAMZA (Diacritical) . used in Urdu
065A:	ﻻ	: ARABIC LETTER BAT

Annotation under

0690:	ﻻ	: ARABIC LETTER DAL with 4 dots above
-------	---	--

. This is not used in Urdu (ref. Page 392 of Unicode 3.0)

Mention of Sindhi in the naming table on pp 392 - 393
under 068E, 06B2, 06B4 should be dropped.

ARABIC -INDIC digits for FOUR, FIVE, SIX and
SEVEN have different glyphs in Sindhi.

These may be added in the annotation

0664	ﻻ	. in Sindhi
0665	ﻻ	. in Sindhi
0666	ﻻ	. in Sindhi
0667	ﻻ	. in Sindhi

Representative Experts:

Dr. M. Hamidullah Bhat, National Council for
Promotion of Urdu Language, MHRD, New Delhi
Tel: 011-6103938 E-mail: urducoun@ndf.vsnl.net.in

Dr. M.J.Warsi, National Council for Promotion of
Urdu Language, New Delhi
Tel:011-6169416 E-mail: urdeducation@hclinfinet.com

Dr. Kishore Vaswani, National Council for Promotion
of Sindhi Language Council, Vadodara
Tel:0265-342246

Shri S.S.Pujari, Coordinator
RC-ILTS, C-DAC, Pune
Tel: 020-5652070 E-mail: pujari@cdac.ernet.in