

IDN TLD

STRING Evaluation

India

APTLD

COLOMBO, SRI LANKA

June 27 -28, 2010

IDN ccTLD request from India

S.No	Language	Script	A-Label	U-Label	Unicode Code Points	String in English	String in Language	ISO 3166
1.	Hindi	Devanagari	xn--h2brj9c	भारत	u+092D, u+093E, u+0930, u+0924	Bharat	भारत	IN
2.	Bangla	Bengali	xn--45brj9c	ভারত	u+09AD, u+09BE, u+09B0, u+09A4	Bharat	ভারত	IN
3.	Telugu	Telugu	xn--fpcrj9c3d	భారత్	u+0C2D, u+0C3E, u+0C30, u+0C24, u+0C4D	Bharat	భారత్	IN
4.	Tamil	Tamil	xn--xkc2dl3a5ee0h	இந்தியா	u+0B87, u+0BA8, u+0BCD, u+0BA4, u+0BBF, u+0BAF, u+0BBE	Indiya	இந்தியா	IN
5.	Gujarati	Gujarati	xn--gecrj9c	ભારત	u+0AAD, u+0ABE, u+0AB0, u+0AA4	Bharat	ભારત	IN
6.	Punjabi/ Panjabi	Gurmukhi	xn--s9brj9c	ਭਾਰਤ	u+0A2D, u+0A3E, u+0A30, u+0A24	Bharat	ਭਾਰਤ	IN
7.	Urdu	Arabic	xn--mgbbh1a71e	بھارت	U+0628, u+06BE, u+0627, u+0631, u+062A	Bharat	بھارت	IN

IDN ccTLD STRING in all 22 official languages of India

	Language	Short form
1.	Assamese	ভাৰত
2.	Bengali	ভাৰত
3.	Bodo	भारत
4.	English	India
5.	Gujarati	ભારત
6.	Hindi	भारत
7.	Kannada	ಭಾರತೀಯ
8.	Kashmiri	بھارت
9.	Konkani	भारत
10.	Malayalam	ഭാരതം
11.	Manipuri (also Meitei or Meithei)	ভাৰত
12.	Marathi	भारत
13.	Nepali	भारत

14.	Oriya	ଭାରତ
15.	Punjabi	ਭਾਰਤ
16.	Sanskrit	भारतम्
17.	Sindhi	ڀارت
18.	Tamil	இந்தியா
19.	Telugu	భారతీయ
20.	Urdu	ہندوستان
21.	Dogri	भारत
22.	Maithili	भारतक
23.	Santhali	भारत

Hindi Language Table

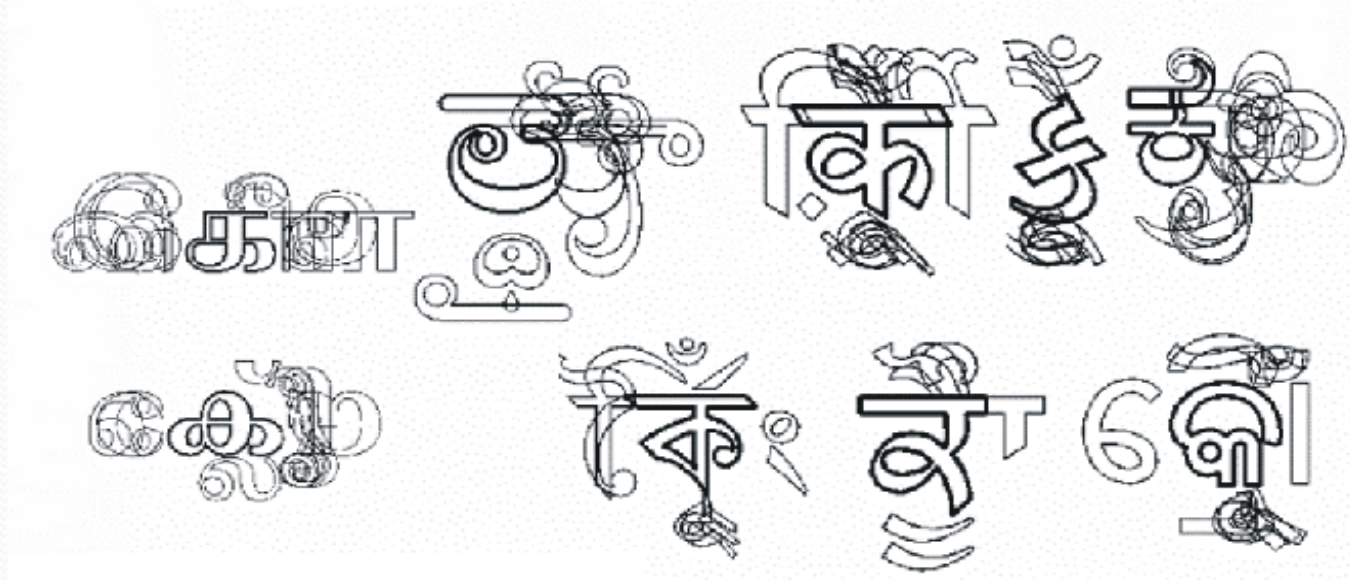
Characters marked in yellow are not applicable to the language.

Devanagari

	090	091	092	093	094	095	096	097
0	ॐ	ऐ 0910	ठ 0920	र 0930	ी 0940	ॐ 0950	ॠ 0960	० 0970
1	ँ 0901	ऑ 0911	ड 0921	ॠ 0931	ॡ 0941	ॢ 0951	ॣ 0961	। 0971
2	ं 0902	ओ 0912	ढ 0922	ल 0932	ॣ 0942	॥ 0952	० 0962	ॠ 0972
3	ः 0903	ओ 0913	ण 0923	ळ 0933	॥ 0943	॥ 0953	॥ 0963	
4	ॐ 0904	औ 0914	त 0924	ळ 0934	॥ 0944	॥ 0954	। 0964	
5	अ 0905	क 0915	थ 0925	व 0935	ँ 0945		॥ 0965	
6	आ 0906	ख 0916	द 0926	श 0936	॥ 0946		० 0966	
7	इ 0907	ग 0917	ध 0927	ष 0937	े 0947		१ 0967	
8	ई 0908	घ 0918	न 0928	स 0938	ै 0948	क 0958	२ 0968	
9	उ 0909	ङ 0919	न 0929	ह 0939	ॉ 0949	ख 0959	३ 0969	
A	ऊ 090A	च 091A	प 092A		ो 094A	ग 095A	४ 096A	
B	ॠ 090B	छ 091B	फ 092B		ो 094B	ज 095B	५ 096B	ग 097B
C	ॡ 090C	ज 091C	ब 092C	ॠ 093C	ौ 094C	ड 095C	६ 096C	ज 097C
D	ँ 090D	झ 091D	भ 092D	ऌ 093D	ॡ 094D	ढ 095D	७ 096D	१ 097D
E	ऐ 090E	ञ 091E	म 092E	ा 093E		फ 095E	८ 096E	इ 097E
	ॠ 090F	ॡ 091F	य 092F	ि 093F		य 095F	९ 096F	४ 097F

Variants

Indian languages being syllabic in nature, the number of visual variations that can be formed with a single character can be numerous.



The number increases manifold when conjuncts come in picture and hence the possibilities of visually similar looking shapes.

Linguistic Challenges we Overcame

- One script used for multiple languages
- One language written in more than one script
- (Homographs) Variants within Same Unicode Page and different Unicode pages
- Written Left to Right as well as Right to Left
- Visually lookalike characters across different scripts
- Intermixing of Languages / Scripts common, especially in names
- Diacritic marks – above and below.
- Web Browsers not yet adaptable to Indian Fonts.

Challenges we Overcame

- Standards
 - Multiple languages from the same code page
 - Universal character coding for IDN use Scripts – Unicode V3.2/V5.0
 - Multiple Input and storage of the same string and therefore rendering (IDN conforming Keyboard)
- Social - Addressing representation of all languages and Dialects
- Legal - Trademarks in Indic Scripts
- Technological – Unorganized Language tool sector, Open-source Solutions

How Did India Arrive at these Strings

Administrative Functions

- Conducted Workshops > 3yrs - Brought together Linguists, Language Technologists, Registrars, Language Communities, TLD Registry, Users, Language Authorities.
- Entrusted .IN Registry to also manage IDN Domain Names through consensus process.
- Backend registration system
- IDN supporting keyboards & Fonts for browsers
- Got affirmations to the TLD string.

Contd...

How Did India Arrive at these Strings

- Followed Universal Character Code
- Extracted Unique Language Character Tables to ensure that only those characters pertinent within a given language are used &
- Script tagging with languages.
- Identified Homographs within Script and across scripts reduce the risk of spoofing and phishing .
- Formulated a generic Augmented Backus Naur Formalism to restrict formation of malformed syllables &
- And Language Specific Policies

Brahmi Syllable Components

Scripts for Indian languages from Brahmi script and have a common phonetic structure, making a common character set possible with some language exceptions.

Consonants(C) : क ख ग घ ङ च छ ज झ ञ ट ठ ड ढ ण त थ द ध न ण
य य र र ल ळ ळ व श ष स ह

Vowels (V) : अ आ इ ई उ ऊ ऋ ऐ ऐ ऋँ ओ ओ औ औ

Matras (M) : ा ि िी ु ू ृ े े ै ॅ ो ो औ ाँ

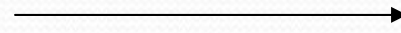
Vowel modifiers (D) : ँ ं ः

Halant (H) : ः

Nukta (N) : ः Operator

Augmented Backus Naur Formalism

Brahmi Syllable
(BNF)



IDN Syllable
(ABNF)

Augmented Backus–Naur Form (ABNF): A Unicode subset, form.

Full-Cons ::= C [N]

Pure-Cons ::= Full-Cons H

Cons-Syllable ::= [Pure-Cons] [Pure-Cons] Pure-Cons

Cons-Vowel-Syllable ::= [Cons-Syllable] Full-Cons [M] [D]

Vowel-Syllable ::= V [D]

Syllable
Syllable ::= Cons-Vowel-Syllable | Vowel-

Word ::= {Syllable} [Cons-Syllable]

ABNF Operators

Operators

`::=` defines a relation.

`{}` enclose items which may be repeated one or more times.

`[]` enclose items which may not be present.

`|` separates items, out of which only one can be present.

ABNF variables against all languages in IDN.

	C	V	M	D	B	X	H	A	N	Y	L	Z
Hindi	✓	✓	✓	✓	✓	✓	✓		✓	✓		
Marathi	✓	✓	✓	✓	✓	✓	✓			✓		
Konkani	✓	✓	✓	✓	✓	✓	✓			✓		
Nepali	✓	✓	✓	✓	✓	✓	✓			✓		
Sanskrit	✓	✓	✓	✓	✓	✓	✓			✓		
Oriya	✓	✓	✓	✓	✓	✓	✓		✓	✓		
Punjabi	✓	✓	✓	✓		✓	✓	✓	✓			
Gujarati	✓	✓	✓	✓	✓	✓	✓			✓		
Bangla	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓
Assamese	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓
Tamil	✓	✓	✓			✓	✓					
Telugu	✓	✓	✓	✓	✓	✓	✓			✓		
Kannada	✓	✓	✓	✓		✓	✓			✓		
Malayalam	✓	✓	✓	✓		✓	✓			✓	✓	

Hindi Language Table

Characters marked in yellow are not applicable to the language.

	090	091	092	093	094	095	096	097
0	ॐ	ऐ 0910	ठ 0920	र 0930	ी 0940	ॐ 0950	ॠ 0960	ॡ 0970
1	ँ 0901	ऑ 0911	ड 0921	ॠ 0931	ॡ 0941	ॢ 0951	ॣ 0961	। 0971
2	ं 0902	ओ 0912	ढ 0922	ल 0932	॥ 0942	० 0952	१ 0962	अँ 0972
3	ः 0903	ओ 0913	ण 0923	ळ 0933	० 0943	१ 0953	२ 0963	
4	ऐ 0904	औ 0914	त 0924	ळ 0934	० 0944	१ 0954	२ 0964	
5	अ 0905	क 0915	थ 0925	व 0935	ँ 0945		॥ 0965	
6	आ 0906	ख 0916	द 0926	श 0936	० 0946		० 0966	
7	इ 0907	ग 0917	घ 0927	ष 0937	े 0947		१ 0967	
8	ई 0908	घ 0918	न 0928	स 0938	ै 0948	ॠ 0958	ॡ 0968	
9	उ 0909	ङ 0919	न 0929	ह 0939	ॉ 0949	ख 0959	३ 0969	
A	ऊ 090A	च 091A	प 092A		ौ 094A	ग 095A	४ 096A	
B	ॠ 090B	ॡ 091B	फ 092B		ो 094B	ज 095B	५ 096B	ग 097B
C	ॢ 090C	ज 091C	ब 092C	ॠ 093C	ौ 094C	ड 095C	६ 096C	ज 097C
D	ँ 090D	झ 091D	भ 092D	ऽ 093D	ॡ 094D	ढ 095D	७ 096D	१ 097D
E	ऐ 090E	ञ 091E	म 092E	ा 093E		फ 095E	८ 096E	इ 097E
F	ए 090F	ट 091F	य 092F	ि 093F		य 095F	९ 096F	ब 097F

Hindi Variant Table

Nukta Variants:

क 0915+093C	क 0958
ख 0916+093C	ख 0959
ग 0917+093C	ग 095A
ज 091C+093C	ज 095B
ड 0921+093C	ड 095C
ढ 0922+093C	ढ 095D
फ 092B+093C	फ 095E

Hindi Variant Table

द्र 0926+094D+0917	द्र 0926+094D+0930	द्र 0926+094D+0928
द्व 0926+094D+0927	द्व 0926+094D+0918	
ष्ट 0937+094D+091F	ष्ट 0937+094D+0920	
क्ष 0936+094D+0935	क्ष 0936+094D+0930+094D+0935	
क्ष 0936+094D+0928	क्ष 0936+094D+0930+094D+0928	

श्च 0936+094D+091A	श्च 0936+094D+0930+094D+091A
क्ष 0936+094D+0932	क्ष 0936+094D+0930+094D+0932
त् 0924+094D+0924+0 94D	त् 0924+094D
ँ 0901	ँ 0945+0902
द्व 0926+094D+0935	द्व 0926+094D+092C

Marathi : Variant Table

द्र 0926+094D+0917	द्र 0926+094D+0930	द्र 0926+094D+0928	क्ष 0936+094D+0932	क्ष 0936+094D+0930+094D+0932
द्व 0926+094D+0927	द्व 0926+094D+0918		त् 0924+094D+0924+094 D	त् 0924+094D
ष्ट 0937+094D+091F	ष्ट 0937+094D+0920		ँ 0901	ँ 0945+0902
क्ष 0936+094D+0935	क्ष 0936+094D+0930+094D+0935		द्व 0926+094D+0935	द्व 0926+094D+092C
क्ष 0936+094D+0928	क्ष 0936+094D+0930+094D+0928		र 0931+ 094D	- 002D
क्ष 0936+094D+091A	क्ष 0936+094D+0930+094D+091A			

Hindi Effective ABNF

Consonant-Syllable → *3(C[N]H) C[N] [H|D|B|X|M[D|B|X]]

Vowel-Syllable → V[D|B|X]

Syllable → Consonant-Syllable [Y] | Vowel-Syllable[Y]

IDN-Label → (Syllable | digit)*([dash](Syllable | digit))

IDN Policy – Salient Features

▣ Length

- Domain name should not be less than X native characters, and the associated Punycode cannot be longer than 63 LDH characters.
- In the IDN .IN, names must be at least 3 ASCII characters long.

▣ Validity of domain String

- Homographs only out of conjuncts and not pure consonants
- Script mixing not allowed.
- Symbols and Abbreviations and punctuation mark characters not allowed **Isshar (√), Abbreviation sign (.), | |**
- IDNs must have at least one non-LDH character with an IDN-Tag.

IDN Policy – Salient Features

- Zero width joiner/Zero Width non Joiner not allowed महाराष्ट्र without ZWJ and ZWNJ

महाराष्ट्र with zero width joiner after हा

महाराष्ट्र with zero width non-joiner after म

- LDH with only Numerals (0 – 9) no Indian Script numerals ०१२३४५६७८९.

▣ String Registration

- Registry requires the registrant to announce the intended language / script of the domain registered.
- To change a language tag between two languages (IDN-Tag change), the domain must be valid in the new tag.

Sample Web Browser transparent Font

Sakal Bharat **Assamese** Normal

কঁহুৱা বন মোৰ অশান্ত মন,

Sakal Bharat **Bengali** Normal

সংস্বেৰ পাকা খাতায় যাঁৱা নাম লেখাতে চান

Sakal Bharat **Devanagari** Normal

भारतीय भाषाएं और सामाजिक परिवर्तन ।

Sakal Bharat **Gujarati** Normal

ભારતીય ભાષાઓ અને સામાજિક પરિવર્તન ।

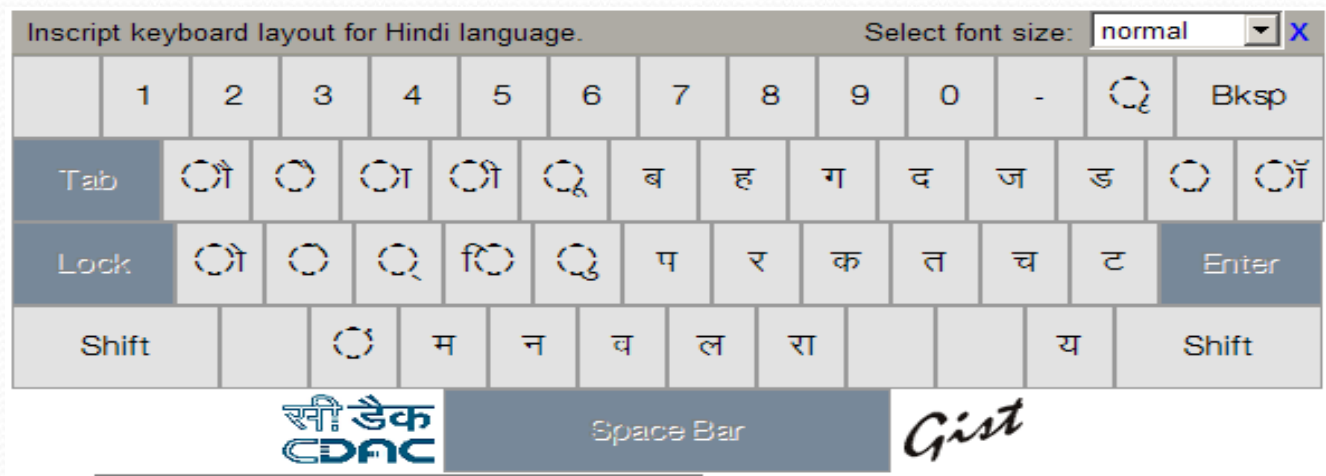
Sakal Bharat **Kannada** Normal

ಇಡೀ ಪ್ರಪಂಚದ ಜನರು ಒಂದೇ ಭಾಷೆಯನ್ನು

FLOATING KEYBOARD-3

Hindi Keyboard

INDIA MILESTONES



Hindi Keyboard shift position:



OS and Browser adoption of Indian IDN string

Sr.no	Operating System	Browsers	Remarks
1	Windows 2000	IE 6.0	Does not support conversion to Punycode by default.
2	Windows 2000	Mozilla 3.0.11	No other major issues found
3	Window XP	IE 6.0	Not working
4	Window XP	Mozilla 3.0.3	No other major issues found
5	Window XP	Chrome	No other major issues found
6	Window XP	Safari	No other major issues found
7	Window XP	Opera 9.64	No other major issues found

8	Window XP	IE 8.0	No other major issues found
9	Windows Vista	IE 7.0	Font's size is very small
10	Windows Vista	Mozilla 3.0.3	No other major issues found
11	Windows Vista	Chrome	No other major issues found
12	Windows Vista	Opera 9.64	No other major issues found
13	Windows Vista	Safari	No other major issues found
14	Ubuntu 8.04	Mozilla 3.0.3	No other major issues found
15	Fedora 6	Mozilla 3.0.3	No other major issues found

Indian languages shared with other Countries (Ethnologue 2009)

S.No	Language	Total Population of Speakers in the world	Countries
1.	Assamese (2)	16,818,750	Bhutan(109,000)
2.	Bengali (6)	181,272,900	Bangladesh, Nepal(23,600), Saudi Arabia(15,000), Canada, United Kingdom(400,000)
3.	Bhojpuri (4)	38,546,000	Mauritius, Suriname, Nepal
4.	Gujarati (11)	46,493,990	Oman, Kenya, Malawi , Mozambique, Uganda, Canada, United States, United Kingdom(140,000), South Africa(26,000), Zimbabwe(19,000)
5.	Hindi (20)	181,676,620	Bangladesh(346,000), Belize(8460), Bhutan(31,900), Canada, Germany(24,500), Guyana, Kenya, Mauritius, Nepal(106,000), New Zealand(11,200), Philippines(2,420), Singapore(5000), United Arab Emirates, Yemen, Sierra Leone, Suriname, Trinidad and Tobago, United States, Zambia
6.	Kannada (3)	35,327,600	Canada, United States
7.	Kashmiri (4)	5,640,940	Pakistan, United Kingdom, United States
8.	Marathi (5)	68,061,130	Canada, Israel, Mauritius, United States
9.	Malayalam (11)	35,893,990	Bahrain, Canada, Fiji, Israel, Malaysia (Peninsular), Qatar, Singapore, United Arab Emirates, United Kingdom, United States

Indian languages shared with other Countries (Ethnologue 2009)

S.No	Language	Total Population of Speakers in the world	Countries
10.	Oriya (2)	31,732,500	Bangladesh(32,500)
11.	Panjabi (12)	91,856,600	Afghanistan, Pakistan, United Arab Emirates, Kenya, Libya(54000), Mauritius (26000), Canada, United States, Canada, United Kingdom(6,00,000), Fiji
12.	Santali (4)	6,170,900	Bangladesh, Bhutan, Nepal
13.	Sindhi (4)	21,382,120.	Oman, Philippines(20,000),United Kingdom(25,000)
14.	Tamil (11)	65,675,200	Bahrain(18,000),Qatar(2200),Malaysia(2000), Singapore, Sri Lanka, United Arab Emirates, South Africa, Canada, United States, Fiji
15.	Telugu (6)	69,758,890	Bahrain(13,000), Singapore, Mauritius, Canada, Fiji
16.	Urdu (9)	60,586,800	Saudi Arabia(382,000), Afghanistan, Bahrain(28,800), Pakistan, Oman, United States, Mauritius, Zambia
Tulika Pandey		Deptt, of Information Technology, Govt.of India	

Common Language Country Harmonized Table - Tamil

Sri Lanka		Singapore		Malaysia		India	
U+0B83	TAMIL SIGN VISARGA = aytham	0B83; TAMIL SIGN VISARGA = aytham	U+0B83	TAMIL SIGN VISARGA = aytham	U+0B83	TAMIL SIGN VISARGA = aytham	
U+0B85	TAMIL LETTER A	0B85; TAMIL LETTER A	U+0B85	TAMIL LETTER A	U+0B85	TAMIL LETTER A	
U+0B86	TAMIL LETTER AA	0B86; TAMIL LETTER AA	U+0B86	TAMIL LETTER AA	U+0B86	TAMIL LETTER AA	
U+0B87	TAMIL LETTER I	0B87; TAMIL LETTER I	U+0B87	TAMIL LETTER I	U+0B87	TAMIL LETTER I	
U+0B88	TAMIL LETTER II	0B88; TAMIL LETTER II	U+0B88	TAMIL LETTER II	U+0B88	TAMIL LETTER II	
U+0B89	TAMIL LETTER U	0B89; TAMIL LETTER U	U+0B89	TAMIL LETTER U	U+0B89	TAMIL LETTER U	
U+0B8A	TAMIL LETTER UU	0B8A; TAMIL LETTER UU	U+0B8A	TAMIL LETTER UU	U+0B8A	TAMIL LETTER UU	
U+0B8E	TAMIL LETTER E	0B8E; TAMIL LETTER E	U+0B8E	TAMIL LETTER E	U+0B8E	TAMIL LETTER E	
U+0B8F	TAMIL LETTER EE	0B8F; TAMIL LETTER EE	U+0B8F	TAMIL LETTER EE	U+0B8F	TAMIL LETTER EE	
U+0B90	TAMIL LETTER AI	0B90; TAMIL LETTER AI	U+0B90	TAMIL LETTER AI	U+0B90	TAMIL LETTER AI	
U+0B92	TAMIL LETTER O	0B92; TAMIL LETTER O	U+0B92	TAMIL LETTER O	U+0B92	TAMIL LETTER O	
U+0B93	TAMIL LETTER OO	0B93; TAMIL LETTER OO	U+0B93	TAMIL LETTER OO	U+0B93	TAMIL LETTER OO	
U+0B94	TAMIL LETTER AU	0B94; TAMIL LETTER AU	U+0B94	TAMIL LETTER AU	U+0B94	TAMIL LETTER AU	
U+0B95	TAMIL LETTER KA	0B95; TAMIL LETTER KA	U+0B95	TAMIL LETTER KA	U+0B95	TAMIL LETTER KA	
U+0B99	TAMIL LETTER NGA	0B99; TAMIL LETTER NGA	U+0B99	TAMIL LETTER NGA	U+0B99	TAMIL LETTER NGA	
U+0B9A	TAMIL LETTER CA	0B9A; TAMIL LETTER CA	U+0B9A	TAMIL LETTER CA	U+0B9A	TAMIL LETTER CA	
U+0B9C	TAMIL LETTER JA	0B9C; TAMIL LETTER JA	U+0B9C	TAMIL LETTER JA	U+0B9C	TAMIL LETTER JA	
U+0B9E	TAMIL LETTER NYA	0B9E; TAMIL LETTER NYA	U+0B9E	TAMIL LETTER NYA	U+0B9E	TAMIL LETTER NYA	
U+0B9F	TAMIL LETTER TTA	0B9F; TAMIL LETTER TTA	U+0B9F	TAMIL LETTER TTA	U+0B9F	TAMIL LETTER TTA	
U+0BA3	TAMIL LETTER NNA	0BA3; TAMIL LETTER NNA	U+0BA3	TAMIL LETTER NNA	U+0BA3	TAMIL LETTER NNA	
U+0BA4	TAMIL LETTER TA	0BA4; TAMIL LETTER TA	U+0BA4	TAMIL LETTER TA	U+0BA4	TAMIL LETTER TA	
U+0BA8	TAMIL LETTER NA	0BA8; TAMIL LETTER NA	U+0BA8	TAMIL LETTER NA	U+0BA8	TAMIL LETTER NA	
U+0BA9	TAMIL LETTER NNNA	0BA9; TAMIL LETTER NNNA	U+0BA9	TAMIL LETTER NNNA	U+0BA9	TAMIL LETTER NNNA	
U+0BAA	TAMIL LETTER PA	0BAA; TAMIL LETTER PA	U+0BAA	TAMIL LETTER PA	U+0BAA	TAMIL LETTER PA	
U+0BAE	TAMIL LETTER MA	0BAE; TAMIL LETTER MA	U+0BAE	TAMIL LETTER MA	U+0BAE	TAMIL LETTER MA	
U+0BAF	TAMIL LETTER YA	0BAF; TAMIL LETTER YA	U+0BAF	TAMIL LETTER YA	U+0BAF	TAMIL LETTER YA	
U+0BB0	TAMIL LETTER RA	0BB0; TAMIL LETTER RA	U+0BB0	TAMIL LETTER RA	U+0BB0	TAMIL LETTER RA	
U+0BB1	TAMIL LETTER RRA	0BB1; TAMIL LETTER RRA	U+0BB1	TAMIL LETTER RRA	U+0BB1	TAMIL LETTER RRA	
U+0BB2	TAMIL LETTER LA	0BB2; TAMIL LETTER LA	U+0BB2	TAMIL LETTER LA	U+0BB2	TAMIL LETTER LA	
U+0BB3	TAMIL LETTER LLA	0BB3; TAMIL LETTER LLA	U+0BB3	TAMIL LETTER LLA	U+0BB3	TAMIL LETTER LLA	
U+0BB4	TAMIL LETTER LLLA	0BB4; TAMIL LETTER LLLA	U+0BB4	TAMIL LETTER LLLA	U+0BB4	TAMIL LETTER LLLA	

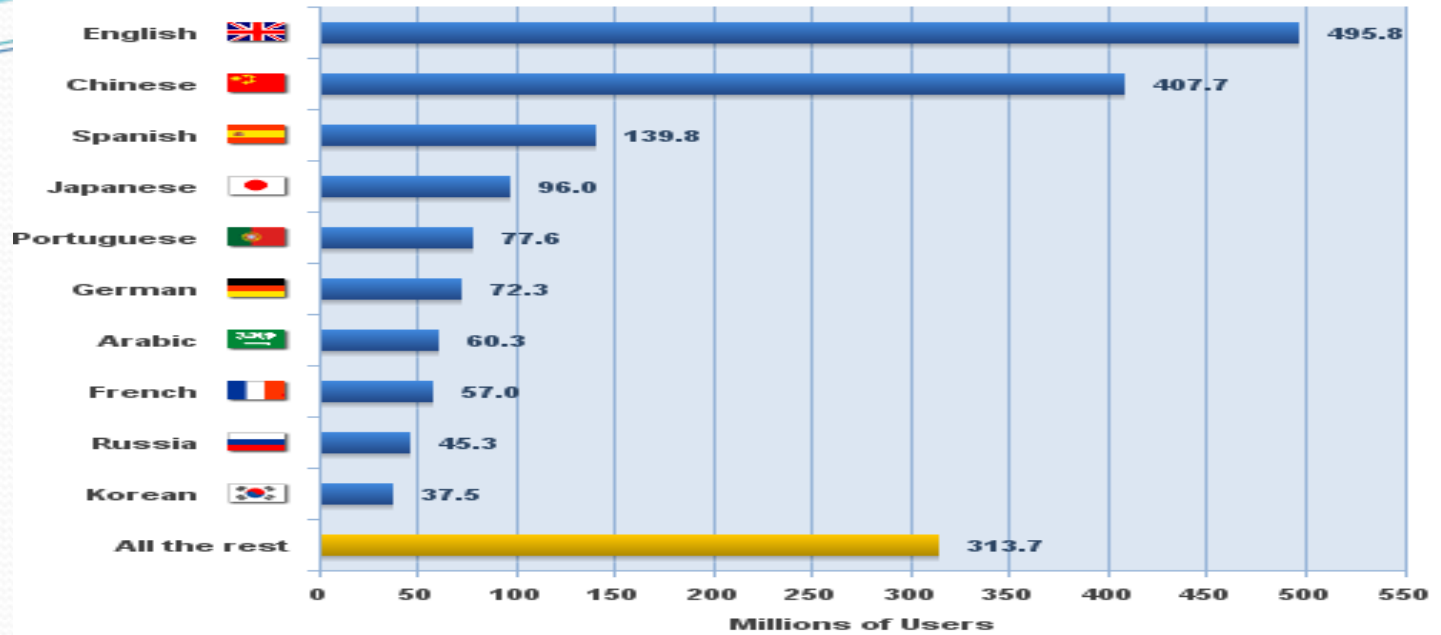
Cross Country Harmonized Table - Tamil

<u>Sri Lanka</u>		<u>Singapore</u>		<u>Malaysia</u>		<u>India</u>	
U+0BB5	TAMIL LETTER VA	0BB5; TAMIL LETTER VA	U+0BB5	TAMIL LETTER VA	U+0BB5	TAMIL LETTER VA	
U+0BB6	TAMIL LETTER SHA		U+0BB6	TAMIL LETTER SHA	U+0BB6	TAMIL LETTER SHA	
U+0BB7	TAMIL LETTER SSA	0BB7; TAMIL LETTER SSA	U+0BB7	TAMIL LETTER SSA	U+0BB7	TAMIL LETTER SSA	
U+0BB8	TAMIL LETTER SA	0BB8; TAMIL LETTER SA	U+0BB8	TAMIL LETTER SA	U+0BB8	TAMIL LETTER SA	
U+0BB9	TAMIL LETTER HA	0BB9; TAMIL LETTER HA	U+0BB9	TAMIL LETTER HA	U+0BB9	TAMIL LETTER HA	
U+0BBE	TAMIL VOWEL SIGN AA	0BBE; TAMIL VOWEL SIGN AA	U+0BBE	TAMIL VOWEL SIGN AA	U+0BBE	TAMIL VOWEL SIGN AA	
U+0BBF	TAMIL VOWEL SIGN I	0BBF; TAMIL VOWEL SIGN I	U+0BBF	TAMIL VOWEL SIGN I	U+0BBF	TAMIL VOWEL SIGN I	
U+0BC0	TAMIL VOWEL SIGN II	0BC0; TAMIL VOWEL SIGN II	U+0BC0	TAMIL VOWEL SIGN II	U+0BC0	TAMIL VOWEL SIGN II	
U+0BC1	TAMIL VOWEL SIGN U	0BC1; TAMIL VOWEL SIGN U	U+0BC1	TAMIL VOWEL SIGN U	U+0BC1	TAMIL VOWEL SIGN U	
U+0BC2	TAMIL VOWEL SIGN UU	0BC2; TAMIL VOWEL SIGN UU	U+0BC2	TAMIL VOWEL SIGN UU	U+0BC2	TAMIL VOWEL SIGN UU	
U+0BC6	TAMIL VOWEL SIGN E	0BC6; TAMIL VOWEL SIGN E	U+0BC6	TAMIL VOWEL SIGN E	U+0BC6	TAMIL VOWEL SIGN E	
U+0BC7	TAMIL VOWEL SIGN EE	0BC7; TAMIL VOWEL SIGN EE	U+0BC7	TAMIL VOWEL SIGN EE	U+0BC7	TAMIL VOWEL SIGN EE	
U+0BC8	TAMIL VOWEL SIGN AI	0BC8; TAMIL VOWEL SIGN AI	U+0BC8	TAMIL VOWEL SIGN AI	U+0BC8	TAMIL VOWEL SIGN AI	
U+0BCA	TAMIL VOWEL SIGN O	0BCA; TAMIL VOWEL SIGN O	U+0BCA	TAMIL VOWEL SIGN O	U+0BCA	TAMIL VOWEL SIGN O	
U+0BCB	TAMIL VOWEL SIGN OO	0BCB; TAMIL VOWEL SIGN OO	U+0BCB	TAMIL VOWEL SIGN OO	U+0BCB	TAMIL VOWEL SIGN OO	
U+0BCC	TAMIL VOWEL SIGN AU	0BCC; TAMIL VOWEL SIGN AU	U+0BCC	TAMIL VOWEL SIGN AU	U+0BCC	TAMIL VOWEL SIGN AU	
U+0BCD	TAMIL SIGN VIRAMA	0BCD; TAMIL SIGN VIRAMA	U+0BCD	TAMIL SIGN VIRAMA	U+0BCD	TAMIL SIGN VIRAMA	
U+0BD0	TAMIL OM	0BD7; TAMIL AU LENGTH MARK					
U+200C	ZERO WIDTH NON-JOINER		U+200C	ZERO WIDTH NON-JOINER			

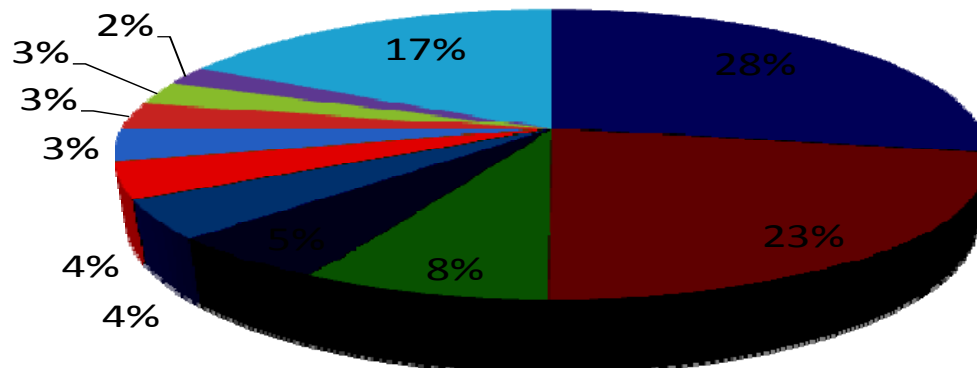
Languages and their Internet Presence

Sl. No.	Continent	Population	Living Languages	No. of Speakers	Internet Presence
1.	Africa	991,002,342	2110	726,453,403	86,217,900
2.	America	927,494,299	993	50,496,321	446,483,050
3.	Asia	3,808,070,503	2322	3,622,771,264	764,435,900
4.	Europe	803,850,858	234	1,553,360,941	425,773,571
5.	Middle East	202,687,005	2322	3,622,771,264	764,435,900
6.	Oceania/ Australia	34,700,201	1250	6,429,788	21,110,490

Top 10 Languages in the Internet in millions of users



Percentage of Top 10 Languages on the Internet



धन्यवाद (Thank You)

Tulika Pandey

Additional Director

tulika@mit.gov.in

<http://www.mit.gov.in>

Department of Information Technology

Ministry of Communications & Information

Technology

Government of India