

**Internationalized Domain Name System  
IDNS Report On Character Constraints In Dzongkha.  
Department Of Information And Technology, 1<sup>st</sup>  
April 2008.**

1. Abstract
2. Dzongkha IDN Table
3. Dzongkha Character Set
4. Basis on which we have chosen a Dzongkha Character to be a valid input for the Dzongkha IDN
5. Algorithm
  - 5.1 Disallowed input
  - 5.2 Valid input
6. Dzongkha Label Separator
7. References

## 1. Abstract

This document provides a report on the character set encoding constraints for IDN in Dzongkha. It gives the characters we have decided as valid input for IDN and which characters have been disallowed. This document also explains how we have come up with this list and on what basis and sorts the problems faced while preparing the Dzongkha character range for IDN.

## 2. Dzongkha IDN Table

Character	Unicode	Description	Dzongkha Table	IDNAbis table
ཨྵ	0F00	TIBETAN SYLLABLE OM	No	Pvalid
ཨྵ	0F01	TIBETAN MARK GTER YIG MGO TRUNCATED A	No	Disallowed
ཨྵ	0F02	TIBETAN MARK GTER YIG MGO - UM RNAM BCAD MA	No	Disallowed
ཨྵ	0F03	TIBETAN MARK GTER YIG MGO - UM GTER TSHEG MA	No	Disallowed
ཨྵ	0F04	TIBETAN MARK INITIAL YIG MGO MDUN MA	No	Disallowed
ཨྵ	0F05	TIBETAN MARK CLOSING YIG MGO SGAB MA	No	Disallowed

འ	0F06	TIBETAN MARK CARET YIG MGO PHUR SHAD MA	No	Disallowed
འ།	0F07	TIBETAN MARK YIG MGO TSHEG SHAD MA	No	Disallowed
འ༎	0F08	TIBETAN MARK SBRUL SHAD	No	Disallowed
འ།།	0F09	TIBETAN MARK BSKUR YIG MGO	No	Disallowed
འ༎།	0F0A	TIBETAN MARK BKA- SHOG YIG MGO	No	Disallowed
་	0F0B	TIBETAN MARK INTERSYLLABIC TSHEG	Yes	Disallowed
་	0F0C	TIBETAN MARK DELIMITER TSHEG BSTAR	Yes	Disallowed
།	0F0D	TIBETAN MARK SHAD	No	Disallowed
།།	0F0E	TIBETAN MARK NYIS SHAD	No	Disallowed
།	0F0F	TIBETAN MARK TSHEG SHAD	No	Disallowed
།	0F10	TIBETAN MARK NYIS TSHEG SHAD	No	Disallowed
།༎	0F11	TIBETAN MARK RIN CHEN SPUNGS SHAD	No	Disallowed
འ༎།	0F12	TIBETAN MARK RGYA GRAM SHAD	No	Disallowed
འ༎།།	0F13	TIBETAN MARK CARET -DZUD RTAGS ME LONG CAN	No	Disallowed

∞	0F14	TIBETAN MARK GTER TSHEG	No	Disallowed
𑍇	0F15	TIBETAN LOGOTYPE SIGN CHAD RTAGS	No	Disallowed
𑍈	0F16	TIBETAN LOGOTYPE SIGN LHAG RTAGS	No	Disallowed
𑍉	0F17	TIBETAN ASTROLOGICAL SIGN SGRA GCAN - CHAR RTAGS	No	Disallowed
𑍊	0F18	TIBETAN ASTROLOGICAL SIGN -KHYUD PA	No	Pvalid
𑍋	0F19	TIBETAN ASTROLOGICAL SIGN SDONG TSHUGS	No	Pvalid
。	0F1A	TIBETAN SIGN RDEL DKAR GCIG	No	Disallowed
∞	0F1B	TIBETAN SIGN RDEL DKAR GNYIS	No	Disallowed
∞	0F1C	TIBETAN SIGN RDEL DKAR GSUM	No	Disallowed
x	0F1D	TIBETAN SIGN RDEL NAG GCIG	No	Disallowed
xx	0F1E	TIBETAN SIGN RDEL NAG GNYIS	No	Disallowed
∞	0F1F	TIBETAN SIGN RDEL DKAR RDEL NAG	No	Disallowed
o	0F20	TIBETAN DIGIT ZERO	Yes	Pvalid

ྱ	0F21	TIBETAN DIGIT ONE	Yes	Pvalid
༢	0F22	TIBETAN DIGIT TWO	Yes	Pvalid
༣	0F23	TIBETAN DIGIT THREE	Yes	Pvalid
༤	0F24	TIBETAN DIGIT FOUR	Yes	Pvalid
༥	0F25	TIBETAN DIGIT FIVE	Yes	Pvalid
༦	0F26	TIBETAN DIGIT SIX	Yes	Pvalid
༧	0F27	TIBETAN DIGIT SEVEN	Yes	Pvalid
༨	0F28	TIBETAN DIGIT EIGHT	Yes	Pvalid
༩	0F29	TIBETAN DIGIT NINE	Yes	Pvalid
ཀ	0F2A	TIBETAN DIGIT HALF ONE	No	Disallowed
ཁ	0F2B	TIBETAN DIGIT HALF TWO	No	Disallowed
ག	0F2C	TIBETAN DIGIT HALF THREE	No	Disallowed
གྷ	0F2D	TIBETAN DIGIT HALF FOUR	No	Disallowed
ང	0F2E	TIBETAN DIGIT HALF FIVE	No	Disallowed
ཅ	0F2F	TIBETAN DIGIT HALF SIX	No	Disallowed
ཆ	0F30	TIBETAN DIGIT HALF SEVEN	No	Disallowed

མ	0F31	TIBETAN DIGIT HALF EIGHT	No	Disallowed
ཉ	0F32	TIBETAN DIGIT HALF NINE	No	Disallowed
ཐ	0F33	TIBETAN DIGIT HALF ZERO	No	Disallowed
ཨ	0F34	TIBETAN MARK BSDUS RTAGS	No	Disallowed
ཨྱ ཨྲ	0F35	TIBETAN MARK NGAS BZUNG NYI ZLA	No	Pvalid
ཨླ	0F36	TIBETAN MARK CARET -DZUD RTAGS BZHI MIG CAN	No	Disallowed
ཨྴ ཨྵ	0F37	TIBETAN MARK NGAS BZUNG SGOR RTAGS	No	Pvalid
ཨྶ	0F38	TIBETAN MARK CHE MGO	No	Disallowed
ཨྷ	0F39	TIBETAN MARK TSA -PHRU	No	Pvalid
ཨྸ	0F3A	TIBETAN MARK GUG RTAGS GYON	No	Disallowed
ཨྐྵ	0F3B	TIBETAN MARK GUG RTAGS GYAS	No	Disallowed
ཨྺ	0F3C	TIBETAN MARK ANG KHANG GYON	No	Disallowed
ཨྻ	0F3D	TIBETAN MARK ANG KHANG GYAS	No	Disallowed
ཨྼ	0F3E	TIBETAN SIGN YAR TSHES	No	Pvalid
ཨ྽	0F3F	TIBETAN SIGN MAR TSHES	No	Pvalid

ཀ	0F40	TIBETAN LETTER KA	Yes	Pvalid
ཁ	0F41	TIBETAN LETTER KHA	Yes	Pvalid
ག	0F42	TIBETAN LETTER GA	Yes	Pvalid
ཁྲ	0F43	TIBETAN LETTER GHA	Yes	Disallowed
ང	0F44	TIBETAN LETTER NGA	Yes	Pvalid
ཅ	0F45	TIBETAN LETTER CA	Yes	Pvalid
ཆ	0F46	TIBETAN LETTER CHA	Yes	Pvalid
ཇ	0F47	TIBETAN LETTER JA	Yes	Pvalid
	0F48	UNASSIGNED		
ཉ	0F49	TIBETAN LETTER NYA	Yes	Pvalid
ཊ	0F4A	TIBETAN LETTER TTA	Yes	Pvalid
ཋ	0F4B	TIBETAN LETTER TTHA	Yes	Pvalid
ཌ	0F4C	TIBETAN LETTER DDA	Yes	Pvalid
ཌྷ	0F4D	TIBETAN LETTER DDHA	Yes	Disallowed
ཎ	0F4E	TIBETAN LETTER NNA	Yes	Pvalid
ཏ	0F4F	TIBETAN LETTER TA	Yes	Pvalid

ཐ	0F50	TIBETAN LETTER THA	Yes	Pvalid
ད	0F51	TIBETAN LETTER DA	Yes	Pvalid
ཏ	0F52	TIBETAN LETTER DHA	Yes	Disallowed
ཎ	0F53	TIBETAN LETTER NA	Yes	Pvalid
པ	0F54	TIBETAN LETTER PA	Yes	Pvalid
ཕ	0F55	TIBETAN LETTER PHA	Yes	Pvalid
བ	0F56	TIBETAN LETTER BA	Yes	Pvalid
བམ	0F57	TIBETAN LETTER BHA	Yes	Disallowed
མ	0F58	TIBETAN LETTER MA	Yes	Pvalid
ཅ	0F59	TIBETAN LETTER TSA	Yes	Pvalid
ཅམ	0F5A	TIBETAN LETTER TSHA	Yes	Pvalid
ཎམ	0F5B	TIBETAN LETTER DZA	Yes	Pvalid
ཎམམ	0F5C	TIBETAN LETTER DZHA	Yes	Disallowed
འ	0F5D	TIBETAN LETTER WA	Yes	Pvalid
འམ	0F5E	TIBETAN LETTER ZHA	Yes	Pvalid
འམམ	0F5F	TIBETAN LETTER ZA	Yes	Pvalid



ཨ	0F60	TIBETAN LETTER - A	Yes	Pvalid
ཨ	0F61	TIBETAN LETTER YA	Yes	Pvalid
ཨ	0F62	TIBETAN LETTER RA	Yes	Pvalid
ཨ	0F63	TIBETAN LETTER LA	Yes	Pvalid
ཨ	0F64	TIBETAN LETTER SHA	Yes	Pvalid
ཨ	0F65	TIBETAN LETTER SSA	Yes	Pvalid
ཨ	0F66	TIBETAN LETTER SA	Yes	Pvalid
ཨ	0F67	TIBETAN LETTER HA	Yes	Pvalid
ཨ	0F68	TIBETAN LETTER A	Yes	Pvalid
ཨ	0F69	TIBETAN LETTER KSSA	Yes	Disallowed
ཨ	0F6A	TIBETAN LETTER FIXED-FORM RA	Yes	Pvalid
	0F6B	UNASSIGNED		
	0F6C	UNASSIGNED		
	0F6D	UNASSIGNED		
	0D6E	UNASSIGNED		
	0F6F	UNASSIGNED		
	0F70	UNASSIGNED		
ཨ	0F71	TIBETAN VOWEL SIGN AA	Yes	Pvalid
ཨ	0F72	TIBETAN VOWEL SIGN I	Yes	Pvalid



འ	0F73	TIBETAN VOWEL SIGN II	Yes	Disallowed
འ	0F74	TIBETAN VOWEL SIGN U	Yes	Pvalid
འ	0F75	TIBETAN VOWEL SIGN UU	Yes	Disallowed
འ	0F76	TIBETAN VOWEL SIGN VOCALIC R	Yes	Disallowed
འ	0F77	TIBETAN VOWEL SIGN VOCALIC RR	Yes	Disallowed
འ	0F78	TIBETAN VOWEL SIGN VOCALIC L	Yes	Disallowed
འ	0F79	TIBETAN VOWEL SIGN VOCALIC LL	Yes	Disallowed
འ	0F7A	TIBETAN VOWEL SIGN E	Yes	Pvalid
འ	0F7B	TIBETAN VOWEL SIGN EE	Yes	Pvalid
འ	0F7C	TIBETAN VOWEL SIGN O	Yes	Pvalid
འ	0F7D	TIBETAN VOWEL SIGN OO	Yes	Pvalid
འ	0F7E	TIBETAN SIGN RJES SU NGA RO	Yes	Pvalid
འ	0F7F	TIBETAN SIGN RNAM BCAD	Yes	Pvalid
འ	0F80	TIBETAN VOWEL SIGN REVERSED I	Yes	Pvalid
འ	0F81	TIBETAN VOWEL SIGN REVERSED II	Yes	Disallowed
འ	0F82	TIBETAN SIGN NYI ZLA NAA DA	No	Pvalid



འཇུག	0F83	TIBETAN SIGN SNA LDAN	No	Pvalid
འཇུག	0F84	TIBETAN MARK HALANTA	No	Pvalid
འཇུག	0F85	TIBETAN MARK PALUTA	No	Disallowed
འཇུག	0F86	TIBETAN SIGN LCI RTAGS	No	Pvalid
འཇུག	0F87	TIBETAN SIGN YANG RTAGS	No	Pvalid
འཇུག	0F88	TIBETAN SIGN LCE TSA CAN	No	Pvalid
འཇུག	0F89	TIBETAN SIGN MCHU CAN	No	Pvalid
འཇུག	0F8A	TIBETAN SIGN GRU CAN RGYINGS	No	Pvalid
འཇུག	0F8B	TIBETAN SIGN GRU MED RGYINGS	No	Pvalid
	0F8C	UNASSIGNED		
	0F8D	UNASSIGNED		
	0F8E	UNASSIGNED		
	0F8F	UNASSIGNED		
འཇུག	0F90	TIBETAN SUBJOINED LETTER KA	Yes	Pvalid
འཇུག	0F91	TIBETAN SUBJOINED LETTER KHA	Yes	Pvalid
འཇུག	0F92	TIBETAN SUBJOINED LETTER GA	Yes	Pvalid
འཇུག	0F93	TIBETAN SUBJOINED	Yes	Disallowed

		LETTER GH A		
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F94	TIBETAN SUBJOINED LETTER NGA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F95	TIBETAN SUBJOINED LETTER CA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F96	TIBETAN SUBJOINED LETTER CHA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F97	TIBETAN SUBJOINED LETTER JA	Yes	Pvalid
	0F98	UNASSIGNED		
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F99	TIBETAN SUBJOINED LETTER NYA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F9A	TIBETAN SUBJOINED LETTER TTA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F9B	TIBETAN SUBJOINED LETTER TTHA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F9C	TIBETAN SUBJOINED LETTER DDA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F9D	TIBETAN SUBJOINED LETTER DDHA	Yes	Disallowed
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F9E	TIBETAN SUBJOINED LETTER NNA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0F9F	TIBETAN SUBJOINED LETTER TA	Yes	Pvalid
ཏཱ་ཅི་ལྷོ་ལྷོ་	0FA0	TIBETAN SUBJOINED	Yes	Pvalid

		LETTER THA		
ཏཱ་	0FA1	TIBETAN SUBJOINED LETTER DA	Yes	Pvalid
ཏཱ་ཨ་	0FA2	TIBETAN SUBJOINED LETTER DHA	Yes	Disallowed
ཏཱ་ཨ་	0FA3	TIBETAN SUBJOINED LETTER NA	Yes	Pvalid
ཏཱ་པ་	0FA4	TIBETAN SUBJOINED LETTER PA	Yes	Pvalid
ཏཱ་ཕ་	0FA5	TIBETAN SUBJOINED LETTER PHA	Yes	Pvalid
ཏཱ་བ་	0FA6	TIBETAN SUBJOINED LETTER BA	Yes	Pvalid
ཏཱ་བ་ཨ་	0FA7	TIBETAN SUBJOINED LETTER BHA	Yes	Disallowed
ཏཱ་མ་	0FA8	TIBETAN SUBJOINED LETTER MA	Yes	Pvalid
ཏཱ་མ་	0FA9	TIBETAN SUBJOINED LETTER TSA	Yes	Pvalid
ཏཱ་མ་ཨ་	0FAA	TIBETAN SUBJOINED LETTER TSHA	Yes	Pvalid
ཏཱ་མ་པ་	0FAB	TIBETAN SUBJOINED LETTER DZA	Yes	Pvalid
ཏཱ་མ་པ་ཨ་	0FAC	TIBETAN SUBJOINED LETTER DZHA	Yes	Disallowed

འ	0FAD	TIBETAN SUBJOINED LETTER WA	Yes	Pvalid
ཅ	0FAE	TIBETAN SUBJOINED LETTER ZHA	Yes	Pvalid
ཎ	0FAF	TIBETAN SUBJOINED LETTER ZA	Yes	Pvalid
ཏ	0FB0	TIBETAN SUBJOINED LETTER -A	Yes	Pvalid
ཐ	0FB1	TIBETAN SUBJOINED LETTER YA	Yes	Pvalid
ཌ	0FB2	TIBETAN SUBJOINED LETTER RA	Yes	Pvalid
ཎ	0FB3	TIBETAN SUBJOINED LETTER LA	Yes	Pvalid
ཏ	0FB4	TIBETAN SUBJOINED LETTER SHA	Yes	Pvalid
ཐ	0FB5	TIBETAN SUBJOINED LETTER SSA	Yes	Pvalid
ཎ	0FB6	TIBETAN SUBJOINED LETTER SA	Yes	Pvalid
ཏ	0FB7	TIBETAN SUBJOINED LETTER HA	Yes	Pvalid
ཐ	0FB8	TIBETAN SUBJOINED LETTER A	Yes	Pvalid
ཎ	0FB9	TIBETAN SUBJOINED	Yes	Disallowed

		LETTER KSSA		
	0FBA	TIBETAN SUBJOINED LETTER FIXED- FORM WA	Yes	Pvalid
	0FBB	TIBETAN SUBJOINED LETTER FIXED- FORM YA	Yes	Pvalid
	0FBC	TIBETAN SUBJOINED LETTER FIXED- FORM RA	Yes	Pvalid
	0FBD	UNASSIGNED		
X	0FBE	TIBETAN KU RU KHA	No	Disallowed
※	0FBF	TIBETAN KU RU KHA BZHI MIG CAN	No	Disallowed
○	0FC0	TIBETAN CANTILLATION SIGN HEAVY BEAT	No	Disallowed
◦	0FC1	TIBETAN CANTILLATION SIGN LIGHT BEAT	No	Disallowed
◐	0FC2	TIBETAN CANTILLATION SIGN CANG TE-U	No	Disallowed
	0FC3	TIBETAN CANTILLATION SIGN SBUB-CHAL	No	Disallowed
	0FC4	TIBETAN SYMBOL DRIL BU	No	Disallowed
	0FC5	TIBETAN SYMBOL RDO RJE	No	Disallowed
	0FC6	TIBETAN SYMBOL PADMA GDAN	No	Pvalid
	0FC7	TIBETAN SYMBOL RDO RJE RGYA GRAM	No	Disallowed
	0FC8	TIBETAN SYMBOL	No	Disallowed

		PHUR PA		
	0FC9	TIBETAN SYMBOL NORBU	No	Disallowed
	0FCA	TIBETAN SYMBOL NORBU NYIS-KHYL	No	Disallowed
	0FCB	TIBETAN SYMBOL NORBU GSUM- KHYIL	No	Disallowed
	0FCC	TIBETAN SYMBOL NORBU BZHI-KHYIL	No	Disallowed
	0FCD	UNASSIGNED	No	
	0FCE	UNASSIGNED	No	
	0FCF	TIBETAN SIGN RDEL NAG GSUM	No	Disallowed
	0FD0	TIBETAN MARK BSKA- SHOG GI MGO RGYAN	No	Disallowed
	0FD1	TIBETAN MARK MNYAM YIG GI MGO RGYAN	No	Disallowed

### **3.Dzongkha Character Set**

Dzongkha character set uses the same character set as the Tibetan. The characters are categorised into consonents, vowels, subjoined consonents, head marks, honorifics, signs and digits.

consonents: There are sixty basic consonents. They make up the basic alphabet.

vowels: These are dependent vowels. They must be used with consonents and subjoined consonents.

subjoined consonents: These are consonents stacked vertically.

Honorifics: These are used while conveying esteem or respect.

tone marks: used to express emphasis.

head marks: used before beginning a letter or a composition.



Therefore it was important to keep in mind the Tibetan IDN which they already have, while preparation of the Dzongkha IDN.

#### **4.Basis on which we have chosen a Dzongkha character to be considered as a valid input for the Dzongkha IDN.**

Frequency of use in writing plain text Dzongkha. We have included all those characters that seem to arise in writing Dzongkha. We have excluded those characters that are complicated and or not used in dzongkha but in Sanskrit.

There was some exception because of irregular writing pattern in Dzongkha for example: The character ka can be vertically stacked where as kha cannot be vertically stacked although the characters representing both case is there. So which means that the character 0F90 (subjoined consonant ka) is valid but 0F91(subjoined character kha) is not. We never have it in writings.

So to simplify things and make a standard we decided to follow the algorithm. Refer below for the Algorithm. All consonants and vowels are included as valid input to the Dzongkha IDN.

There is one thing that deviates from this algorithm. The algorithm doesnot allow punctuation. But with the Dzongkha character 0F0B (Tsheg) and 0F0C (Bstar) which are categorised as punctuation should be allowed. It is because they are syllable marker and words cannot be formed without them.

### **5.Algorithm. (reference from IDNA200x Rationale, internet Draft)**

#### **5.1.Disallowed**

Characters are placed into disallowed if they fall into one or more of the following groups:

- a. The character is a compatibility equivalent for another character, ie application of the normalization method to the character yields some other character.
- b. The character is an upper case form or some other form that is mapped to another character by unicode case folding.
- c. The character is a symbol or punctuation or simply not a letter or a digit.

#### **5.2.Valid**

Characters that are either letter or digit. In some cases those characters that doesnot fall under these categories are also considered valid when they are essential to writing labels and cannot do without them.

Eg: The character 0F0B (Tsheg) and 0F0C (Bstar) are considered punctuation but essential

because they are syllable markers.

### **6.Dzongkha Label Seperator.**

In choosing the label seperator, we had to also consider the Tibetan IDN. Choosing any label seperator would mean disallowing that character for the Tibetans. Since the Tibetans already had their label seperator defined, we decided to use the same. Therefore we are keeping 0F14 (Gter Tsheg) as the label seperator.

Other Things considered were:

1. it should provide clear distinction on its responsibilities. In writings 0F14 (GTER TSHEG) is used as a sentence end marker after TERMA(teasure scripts).
2. Keep Tibetan language in mind too.
3. The label separator should be one specified in the keyboard specification. That is, if a user cannot type in that character, then we might have an unusable label separator.
4. The label seperator should be easy to type that is not in any of the shift state.

### **7.References.**

<http://www.unicode.org/charts/>

Internet-Draft, The unicode Codepoints and IDNA, draft-faltstrom-idnabis-tables-04.txt.

Internet-Draft, Internationalizing Domain Names for Application : Issues, Explanation, and Rationale, draft-klensin-idnabis-issues-07.txt.