

Notes on the Devnagari characters in this dataset

Devnagari is descendent of the ancient Brahmi script and is the script of many languages including Hindi, Marathi, Gujrati, Nepali and Sanskrit. Devnagari is written in non-cursive or discrete form from left to right. Each letter represents a consonant, which carries an inherent vowel *a*. This inherent vowel can be muted by the vowel *halanta*, or can be modified by other vowels. Two or more consonants can be combined together with vowels to form ligatures, or conjuncts or *samyuktasharas*.

As per the Unicode standard 5.1 (based on ISCII 1988), the Devnagari script consists of 112 base symbols including digits, additional consonants and vowels for various other languages like Sanskrit, Marathi, Sindhi other than those used popularly in Hindi.

Following is the Unicode chart for Devnagari script.

	090	091	092	093	094	095	096	097
0	ॐ	ऐ	ठ	र	ी	ॐ	ऋ	०
1	ँ	ऑ	ड	र	ु	ं	ॠ	ॠ
2	ं	ओ	ढ	ल	ॡ	ॢ	ॣ	अँ
3	ः	ओ	ण	ळ	ॣ	॥	॥	
4	॥	औ	त	ळ	॥	॥	॥	
5	अ	क	थ	व	ॠ		॥	
6	आ	ख	द	श	ॠ		०	
7	इ	ग	ध	ष	ॠ		१	
8	ई	घ	न	स	ॠ	ॠ	२	
9	उ	ङ	न	ह	ॠ	ख	३	
A	ऊ	च	प		ॠ	ग	४	
B	ॠ	छ	फ		ॠ	ज	५	ग
C	ॠ	ज	ब	ॠ	ॠ	ड	६	ज
D	ॠ	झ	भ	ॠ	ॠ	ढ	७	१
E	ॠ	ज	म	ॠ		ॠ	८	ड
F	ॠ	ट	य	ॠ		ॠ	९	ब

The dataset collected at HP Labs, India consists of the symbols popularly used to write Hindi. The dataset contains 111 distinct symbols which comprises 36 consonants, 26 half consonants, 12 vowels, 11 vowels in independent form, 13 ligatures (samyuktaksharas) and 11 special symbols. All the symbols are collected with the shirorekha (class id 0) except for the vowels. Samples of two vowels were collected more than once to record different possible positions with respect to the main consonant.

The following table shows the symbols in the dataset.

[illegible]

The following table lists out the symbols present in the dataset along with their character ids and UNICODE symbols.

Character Id.	Symbol	Unicode
0	—	
1	अ	0905
2	आ	0906
3	इ	0907
4	ई	0908
5	उ	0909
6	ऊ	090A
7	ऋ	090B
8	ए	090F
9	ऐ	0910
10	ओ	0913
11	औ	0914
12	क	0915
13	ख	0916
14	ग	0917
15	घ	0918
16	ङ	0919
17	च	091A
18	छ	091B
19	ज	091C
20	झ	091D
21	ञ	091E
22	ट	091F
23	ठ	0920
24	ड	0921
25	ढ	0922
26	ण	0923
27	त	0924
28	थ	0925
29	द	0926

30	ध	0927
31	न	0928
32	प	092A
33	फ	092B
34	ब	092C
35	भ	092D
36	म	092E
37	य	092F
38	र	0930
39	ल	0932
40	व	0935
41	श	0936
42	ष	0937
43	स	0938
44	ह	0939
45	क्ष	0915 094D 0936
46	त्र	0924 094D 0930
47	ज्ञ	
48	ट	093E
49	डि	093F
50	ढी	0940
51	ॢ	0941
52	ॣ	0942
53	।	0943
54	॥	0947
55	०	0948
56	१	094B
57	२	094C
58	३	0902
59	४	0945
60	५	0901
61	६	
62	७	094D
63	८	093C
64	९	0964

65	व	0915 094D
66	ख	0916 094D
67	ग	0917 094D
68	घ	0918 094D
69	च	091A 094D
70	ज	091C 094D
71	झ	091D 094D
72	ञ	091E 094D
73	ट	0923 094D
74	ठ	0924 094D
75	ड	0924 094D 0924 094D
76	ण	0925 094D
77	त	0927 094D
78	थ	0928 094D
79	द	092A 094D
80	ध	092B 094D
81	न	092C 094D
82	प	092D 094D
83	फ	092E 094D
84	ब	092F 094D
85	।	
86	।	
87	।	
88	ल	0932 094D
89	ळ	0935 094D
90	श	0936 094D
91	ष	0937 094D
92	स	0938 094D
93	ह	0939 094D
94	क्ष	0915 094D 0936
95	श्च	0936 094D 0930 094D
96	क्त	0915 094D 0924
97	द्ध	091F 094D 091F
98	त्त	0924 094D 0924
99	द्द	0926 094D 0926

100	द्व	0926 094D 0935
101	घ	0926 094D 092F
102	ख	0926 094D 0927
103	न	0928 094D 0928
104	श	0936 094D 0930
105	हम	0939 094D 092E
106	ह्य	0939 094D 092F
107	ह	0939 094D 0930
108	हृ	0939 094D 0943
109	रु	Matra Only
110	रु	Matra Only