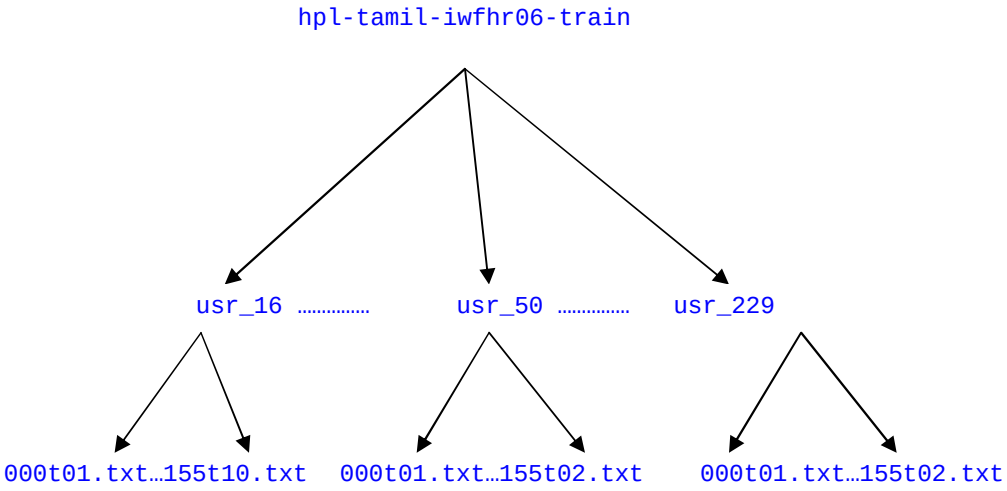
IWFHR06 Tamil Handwritten Character Recognition Competition

Training Dataset *hpl-tamil-iwfhr06-train*

Updated Jan 25, 2006

The training dataset hpl-tamil-iwfhr06-train contains samples of the 156 character classes collected from different writers using a TabletPC application. Most writers contributed two samples per class ("trials"), a few contributed as many as 10. Details of the directory structure, file contents and statistics regarding the number of samples per class are provided in the following sections.

1. Directory Structure



- *hpl-tamil-iwfhr06-train* directory is the root directory.
- Ink data is organized by writer into subdirectories of the form usr_<*user-id*>, e.g., usr_16 corresponds to the 16th writer.
  - User-ids are in the range 16 … 229, but not contiguous
- Ink data is stored in files of the form <*3-digit class-id*>t<*trial-id*>, e.g. 008t03.txt implies the 3rd trial of the character with class-id 008.
  - Class-id is in the range 000 …155
  - Trial-id is in the range 01 … 02 for most users, 01 … 10 for some. However they are not guaranteed to be contiguous since bad samples may have been removed.

2. Ink File Contents

- In each file, ink data is represented in UNIPEN v1.0 format, as shown below.
  - The channels reported for each ink point are X,Yand T. Files corresponding to some users have valid T (time) values for the first and last points of each stroke, with intermediate values set to 0. For other users, the time channel is set to 0 for all points.
  - Since the ink was captured using Microsoft Ink Picture Controls on a TabletPC, spatial resolution and sampling rate correspond to the interpolated ink returned by the control, and are higher than what is supported by the hardware.

```
.VERSION 1.0
.HIERARCHY CHARACTER
.COORD X Y T
.SEGMENT CHARACTER
.X_POINTS_PER_INCH 2500
.Y_POINTS_PER_INCH 2500
.POINTS_PER_SECOND 1200
.PEN_DOWN
935 523 0
935 523 0
935 523 0
935 520 0
935 517 0
935 514 0
935 511 0
.PEN_UP
```

For more information on UNIPEN format please refer to http://unipen.nici.ru.nl/unipen.def

3. Samples per class

Character Id	No of Samples
0	347
1	341
2	340
3	332
4	339
5	345
6	341
7	342
8	334
9	329
10	324
11	340
12	337
13	331
14	331
15	333
16	341
17	341
18	332
19	326
20	339
21	339
22	337
23	326
24	335
25	337
26	332
27	325
28	335

29	323
30	169
31	312
32	344
33	333
34	299
35	334
36	331
37	337
38	335
39	338
40	338
41	339
42	329
43	339
44	331
45	336
46	338
47	338
48	339
49	338
50	339
51	340
52	342
53	335
54	325
55	169
56	331
57	313
58	342
59	314
60	334
61	342
62	337
63	343
64	336
65	336
66	339
67	332
68	336
69	331
70	338
71	339
72	330
73	330
74	329
75	331
76	336
77	335
78	169

79	335
80	331
81	337
82	336
83	334
84	336
85	336
86	337
87	342
88	339
89	332
90	335
91	336
92	338
93	333
94	338
95	326
96	325
97	331
98	337
99	322
100	310
101	323
102	327
103	338
104	330
105	343
106	343
107	340
108	341
109	339
110	343
111	329
112	340
113	277
114	330
115	331
116	331
117	329
118	347
119	329
120	337
121	316
122	286
123	274
124	168
125	304
126	291
127	293
128	282

129	162
130	283
131	280
132	336
133	324
134	327
135	337
136	342
137	342
138	344
139	341
140	333
141	339
142	342
143	328
144	343
145	339
146	341
147	333
148	339
149	338
150	336
151	315
152	170
153	330
154	304
155	344