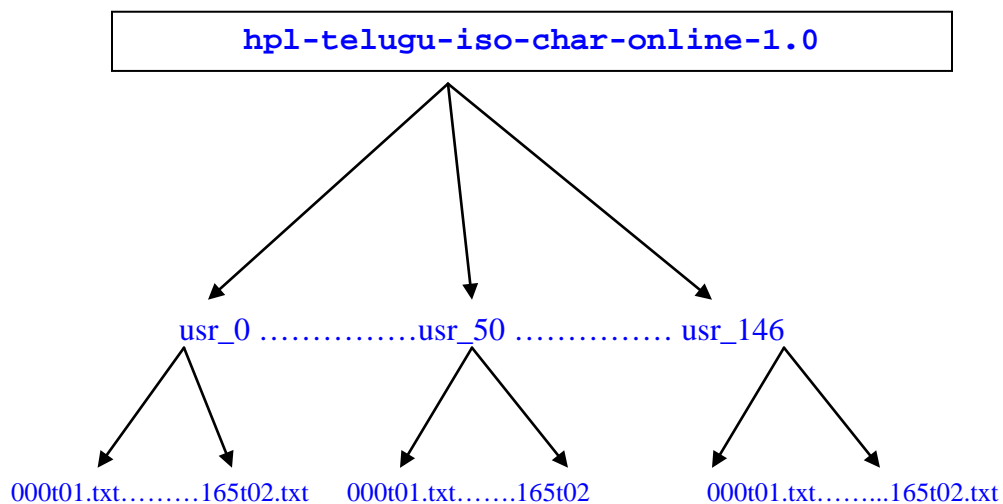


## Dataset *hpl-telugu-iso-char-online-1.0*

Updated Jun 9, 2009

The dataset *hpl-telugu-iso-char-online-1.0* contains samples of the 166 character classes collected from different writers on ACECAD DigiMemo (A4 sized) using a *AcecadDigiMemoDCT* application. The data collection forms were designed using the Adobe Designer, and were printed on A4 sheets. The writers were asked to write on the printed A4 sheets using ACECAD DigiMemo. Most writers contributed two samples per class (“trials”). 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-telugu-iso-char-online-1.0* directory is the root directory for the online data.
- Ink data is organized by writer into subdirectories of the form *usr\_<user-id>*, e.g., *usr\_16* corresponds to the 17<sup>th</sup> writer.
  - User-ids are in the range 0...146, but not contiguous. Samples of four writers (*usr\_24*, *usr\_58*, *usr\_93*, and *usr\_97*) were found to be very noisy and were discarded.
- Ink data is stored in files of the form *<3-digit class-id>t<trial-id>*, e.g. *008t03.txt* implies the 3<sup>rd</sup> trial of the character with class-id 008.
  - Class-id is in the range 000 ...165
  - Trial-id is in the range 01 ... 02. However, some samples written by a user might have been found noisy and hence discarded.

### 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,Y and 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.

```

.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.	Number of Samples
0	267
1	279
2	282
3	271
4	260
5	258
6	263
7	269
8	244
9	266
10	271
11	281
12	272
13	265
14	272
15	276
16	265
17	276
18	279
19	268
20	269
21	274
22	262
23	278
24	274
25	269
26	274
27	280
28	279

29	264
30	272
31	278
32	272
33	282
34	278
35	277
36	271
37	270
38	314
39	278
40	270
41	267
42	276
43	280
44	282
45	271
46	266
47	278
48	278
49	277
50	270
51	274
52	273
53	276
54	277
55	279
56	279
57	280
58	277
59	278
60	270
61	277
62	272
63	270
64	268
65	266
66	283
67	257
68	268
69	272
70	270
71	291
72	257
73	276
74	260
75	296
76	270
77	273
78	280

<b>79</b>	272
<b>80</b>	276
<b>81</b>	262
<b>82</b>	261
<b>83</b>	274
<b>84</b>	275
<b>85</b>	279
<b>86</b>	265
<b>87</b>	287
<b>88</b>	270
<b>89</b>	276
<b>90</b>	270
<b>91</b>	269
<b>92</b>	262
<b>93</b>	257
<b>94</b>	277
<b>95</b>	287
<b>96</b>	271
<b>97</b>	261
<b>98</b>	273
<b>99</b>	265
<b>100</b>	269
<b>101</b>	266
<b>102</b>	264
<b>103</b>	276
<b>104</b>	268
<b>105</b>	272
<b>106</b>	269
<b>107</b>	277
<b>108</b>	279
<b>109</b>	272
<b>110</b>	270
<b>111</b>	265
<b>112</b>	264
<b>113</b>	260
<b>114</b>	282
<b>115</b>	273
<b>116</b>	276
<b>117</b>	273
<b>118</b>	265
<b>119</b>	274
<b>120</b>	283
<b>121</b>	277
<b>122</b>	268
<b>123</b>	277
<b>124</b>	280
<b>125</b>	269
<b>126</b>	280
<b>127</b>	272
<b>128</b>	260

<b>129</b>	292
<b>130</b>	269
<b>131</b>	268
<b>132</b>	262
<b>133</b>	274
<b>134</b>	274
<b>135</b>	278
<b>136</b>	274
<b>137</b>	267
<b>138</b>	307
<b>139</b>	271
<b>140</b>	263
<b>141</b>	239
<b>142</b>	274
<b>143</b>	260
<b>144</b>	268
<b>145</b>	281
<b>146</b>	270
<b>147</b>	278
<b>148</b>	277
<b>149</b>	282
<b>150</b>	273
<b>151</b>	276
<b>152</b>	265
<b>153</b>	266
<b>154</b>	287
<b>155</b>	284
<b>156</b>	277
<b>157</b>	264
<b>158</b>	272
<b>159</b>	266
<b>160</b>	275
<b>161</b>	279
<b>162</b>	263
<b>163</b>	265
<b>164</b>	267
<b>165</b>	255