Meaning of the fields in each line of training and test data:

1. id: each letter is assigned a unique integer id

2. letter: a-z

3. next\_id: id for next letter in the word, -1 if last letter

4. word\_id: each word is assigned a unique integer id

5. position: position of letter in the word

6. p\_{ij}: 0/1-value of pixels in row i, column j (repeat for 16\*8=128 pixels)

Some statistics:

train.txt:

3438 words

25953 letters

test.txt:

3439 words

26198 letters