L2X results report – johmy592, erino397

Problem 1:
After reading the character-map into a dictionary we can lookup the pronouciation 'yi' and we found 484 characters associated with that pronounciation.

Problem 2:
We have implemented a recursive n-gram model with Witten-Bell smoothing. The order can be specified by the user, in our examples we found that orders greater than 4 no longer prvided any increase in accuracy, therefore order 4 was used in the examples below. Order 4 means that our model calculates probabilities based on n-grams with n = 1,2,3,4 . See comments in the jupyter notebook file for implementation details.

Problem 3:
Here are the 10 first predictions made by the model with order 4 along with their probabilities:

(1) Prediction: 想, Prob: 0.00093799
(2) Prediction: 通, Prob: 0.00055108
(3) Prediction: 的, Prob: 0.02013015
(4) Prediction: 价, Prob: 0.00055436
(5) Prediction: 格, Prob: 0.43585072
(6) Prediction: ，, Prob: - (1.0)
(7) Prediction: 可, Prob: 0.00276513
(8) Prediction: 以, Prob: 0.46028784
(9) Prediction: 安, Prob: 0.00141037
(10) Prediction: 装, Prob: 0.13640034

Problem 4:
The model achieves an accuracy of 81.794% on the full test-set.