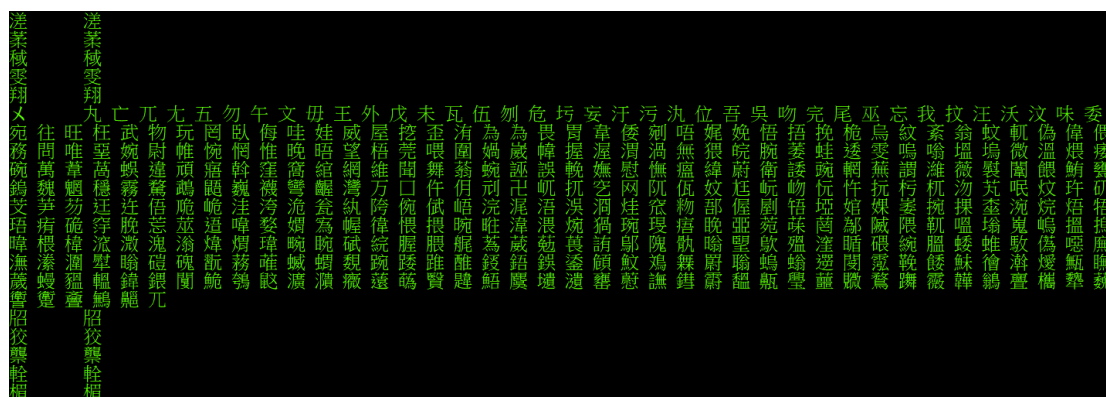


B05902049 資工三 簡崇安

Execute : make map(產生 ZhuYin-Big5.map) / make run(產生 result2/1.txt~10.txt)

mapping.py:利用 python 中的 dict()建立一個 dictionary 以注音以及國字為 index, 如果是注音後面就 append 以此注音為頭的國字,而國字則接著本身,最後再將 dictionary 依序寫入 ZhuYin-Big5.map,產生結果如下



mydisambig.cpp:讀入 bigram.lm, ZhuYin-Big5.map 以及 segment 過的 txt 檔,利用動態規劃來計算所有可能中機率最大者,如果在 txt segment 有國字的便只有一個 candidate,直接得到這個字的 index 即可,否則要從 map 中拿到所有可能的字來計算個別的機率並取最大值,而機率最大的 candidate 便存入 backtrack,最後走過一次 backtrack 就可以填滿本來的句子了,產生結果如下

[illegible]