DSP hw3 report

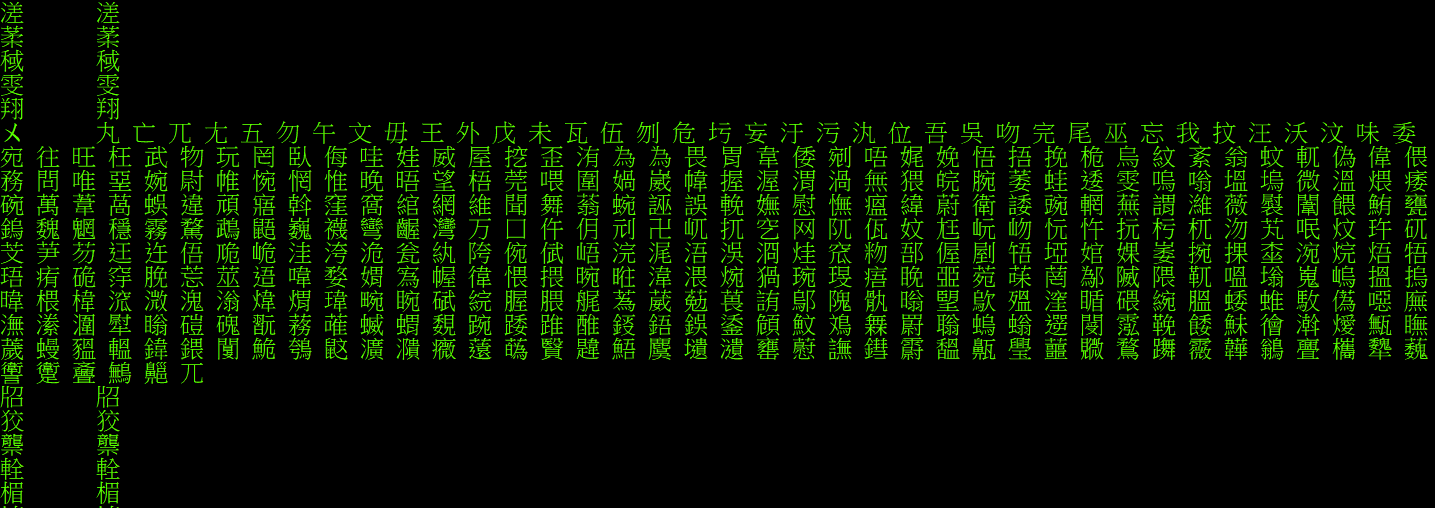
B05902049 資工三 簡崇安

Environment : Win10下的bash

Compile : make all(編譯mydisambig.cpp)

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

