

B05902037 資工三 朱玟嶸

環境：資工系工作站

compile:依據助教在投影片的內容，依序

make all

make map

make run

即可

execute:

make run

usage: (./mydisambig ZhuYin-Big5.map bigram.lm 2 [input_file] > \$output)

先用 mapping 把 Big5-ZhuYin.map ZhuYin-Big5.map。

在 mydisambig 中，先讀入 ZhuYin-Big5.map，因為每個中文字都是兩個 char，所以用 uint16_t 來儲存一個中文字，建立一個 map< uint16_t, vector<uint16_t>> 來儲存 ZhuYin-Big5.map，接著使用 srilm 的 library，除存 language model，接著讀入 query，運用 viterbi 計算出最佳的路徑。

Viterbi:

用兩個 vector 來儲存，一個儲存前一個字所有可能 word 的機率，後面一個儲存當前的最佳機率，當讀入下一個 word，就用 bigram 讀取它跟前一個所有可能的字的機率，以此算出最佳路徑，到最後一個字時，則把空字當作其後面一個字計算機率，最後用 backtrack 來重建出最佳的路徑。

心得：在實做的過程中，被編碼格式拖延了許久，對於中文的編碼處理仍要加強。