

(1) What is your environment?

Macosx

(2) How to “make” your program?

make map (產生 ZhuYin-Big5.map)

make all (產生所要的 mydisambig)

make run (執行 mydisambig 翻譯 1.txt~10.txt)

(3) How to “execute” your program?

第一部分：在 make map 就可以產生 ZhuYin-Big5.map 了

在 result1 資料夾下可以看到 1.txt~10.txt

第二部分：在 make run 就可以沒有 error 執行翻譯工作。(在 make run 之前要先 make all)

在 result2 資料夾下可以看到 1.txt~10.txt

(4) What have you done?

第一部分：

用 python 寫了 mapping.py 的程式，根據 Big5-ZhuYin.map。

裡面用了 split 切割 Big5-ZhuYin.map，也用了 dict 資料結構去存所要對應的資料。用 csv.writer 輸出成 ZhuYin-Big5.map。

第二部分：

用 c++寫了 mydisambig.cpp 的程式。

讀檔寫了大部份：

用到的資料結構 map，set，string，vector。

Map

用 fgets 拿 ZhuYin-Big5.map，用 strtok 切割 data，

用 map<string ,set<string>>mapping 資料結構去存所要對應的資料。

1.txt~10.txt

用 fget 拿 1.txt~10.txt，用 strtok 切割 data，

用 vector<string> splitLine 資料結構去存所要的資料。

Viterbi

用 mapping 找出 splitLine 每個注音文的所有可能，藉由 SRILM 的 disambig 算出 bigram 最大的可能，最後最好的 path 存在 vector<string>BestLine。