

在csie workstation上實作

首先執行**make clean**把不必要的檔案先清掉

接著打**make map**把我們需要的map file建出來

make map 會執行 python3 mapping.py這行指令

mapping是用python3做的

實做的方法為先一行一行讀進原來的big5-ZhuYin.map，用一個Dictionary去紀錄ㄣㄣㄣ等注音，再把對應的國字放入注音那格中

完成mapping後執行**make** 將mydisambig這個執行檔建出來

實作mydisambig的方法是先現成的函式讀進bigram的資料，接著用c++的map讀進剛剛做好的map。讀入map要注意由於他有空白所以要做一些分割的處理。之後就是讀進我們要判斷的資料同樣是一行一行讀進並把空白的字元處理掉。接下來用一個二維的陣列去存我們讀進來的資料有什麼對應的可能。像是讀進的注音去map中找所有那個注音有可能的字並放入陣列中。值得注意的是開頭要放入<s>結尾要加入</s>。最後就是我們的viterbi algorithm，我們一個字一個字去找，去看說每一個接下來的字當中，在他的所有可能當中最有可能是來自前一個字的哪種可能，這些就是bigram中的資料。跑完後在回頭看應該印出什麼句子。

在最後執行**make run**，他會執行像這樣的指令./mydisambig -lm bigram.lm -map ZhuYin-Big5.map -text testdata/1.txt > result2/1.txt1。結果會輸出至result2資料夾中對應的檔案。