

## DSP HW3

(1)What is your environment?

答：64bit workstation in CSIE

(2)How to “make” your program?

答：

i.在目錄底下輸入make指令，此處範例採投影片助教路徑：

```
make MACHINE_TYPE=i686-m64 SRIPATH=/home/ta/srilm-1.5.10 all
```

ii.會產生mapping與mydisambig檔案

iii.在寫作業的平台上，是i686-m64，CSIE的工作站

(3)How to “execute” your program?

答：

a.輸入make map，便可以藉由Big5-ZhuYin.map產生ZhuYin-Big5.map

b.最後，輸入make run，若以助教投影片路徑指令則為：

```
make MACHINE_TYPE=i686-m64 SRIPATH=/home/ta/srilm-1.5.10 LM=bigram.lm run
```

，產生所有結果，並儲存到./result2中

(4)What have you done?

答：

a.在CSIE工作站中架設好了環境，修改Makefile中的路徑與common/Makefile.machine.i686-m64中的幾行程式碼，再用make MACHINE\_TYPE=i686-m64 World指令compile出執行檔。

b.用perl separator\_big5.pl corpus.txt >corpus\_seg.txt指令，將corpus.txt與所有測資檔案中的一個一個單字都分隔開來

c.執行./ngram-count -read lm.cnt -lm bigram.lm -unk -order 2與./ngram-count -read lm.cnt -lm bigram.lm -unk -order 2後，便可得出bigram.lm檔案

d.寫出mapping檔，藉由Big5-ZhuYin.map產生ZhuYin-Big5.map，可以將注音對照到其注音作開頭的所有國字（此處會額外用到zy\_table.txt檔來讀取注音）

e.分為兩個部分：

i.直接用tool，以SRILM內的工具，執行./disambig -text ./testdata/seg\_i.txt -map ZhuYin-Big5.map -lm bigram.lm -order 2 > ./result1/i.txt，將1~10.txt testing data中的注音文轉換為中文，並化為整串句子

ii.寫出mydisambig，可以將所有testing data中的注音文，藉由viterbi演算法，再加上ZhuYin-Big5.map、bigram.lm、10筆測資，來把注音文轉換為中文