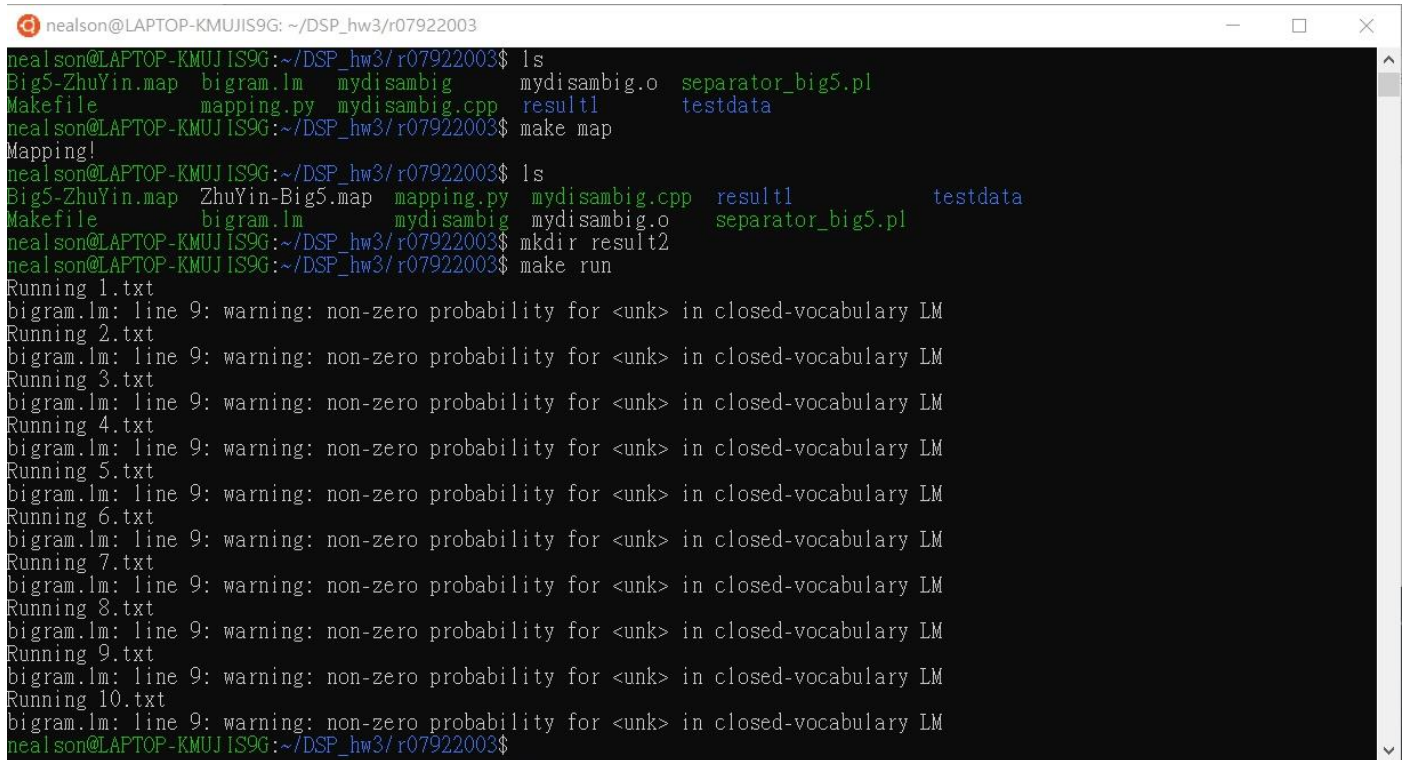


DSP HW3 資工碩一 r07922003 劉濬慶

本次作業 mapping 用 python 寫的 mydisambig 用 c++ 寫的

1. 執行環境: Win10 內建下的虛擬 ubuntu bash



```
nealson@LAPTOP-KMUJIS9G: ~/DSP_hw3/r07922003$ ls
Big5-ZhuYin.map  bigram.lm  mydisambig  mydisambig.o  separator_big5.pl
Makefile         mapping.py  mydisambig.cpp  result1      testdata
nealson@LAPTOP-KMUJIS9G:~/DSP_hw3/r07922003$ make map
Mapping!
nealson@LAPTOP-KMUJIS9G:~/DSP_hw3/r07922003$ ls
Big5-ZhuYin.map  ZhuYin-Big5.map  mapping.py  mydisambig.cpp  result1      testdata
Makefile         bigram.lm        mydisambig  mydisambig.o    separator_big5.pl
nealson@LAPTOP-KMUJIS9G:~/DSP_hw3/r07922003$ mkdir result2
nealson@LAPTOP-KMUJIS9G:~/DSP_hw3/r07922003$ make run
Running 1.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 2.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 3.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 4.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 5.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 6.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 7.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 8.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 9.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
Running 10.txt
bigram.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM
nealson@LAPTOP-KMUJIS9G:~/DSP_hw3/r07922003$
```

2. 執行步驟

Step1: make map 會將 Big5-ZhuYin.map -> ZhuYin-Big5.map

Step2: make 編譯 mydisambig.cpp

Step3: mkdir result2 建立資料夾(此步驟在 test_env 可省略)

Step4: make run 會將 output 放入 result2 中(改成以下的指令)

```
./mydisambig -text testdata/$$i.txt -map $(TO) -lm $(LM) -order 2 result2/$$i.txt; \
```

3. 前處理

安裝 srilm-1.5.10，因為是 win10-64 bit 所以安裝時用 Makefile.machine.i686-m64，依步驟切割 1~10.txt 的 segmentation 檔案以及生成 bigram.lm

4. What you have done

1. 安裝 srilm-1.5.10

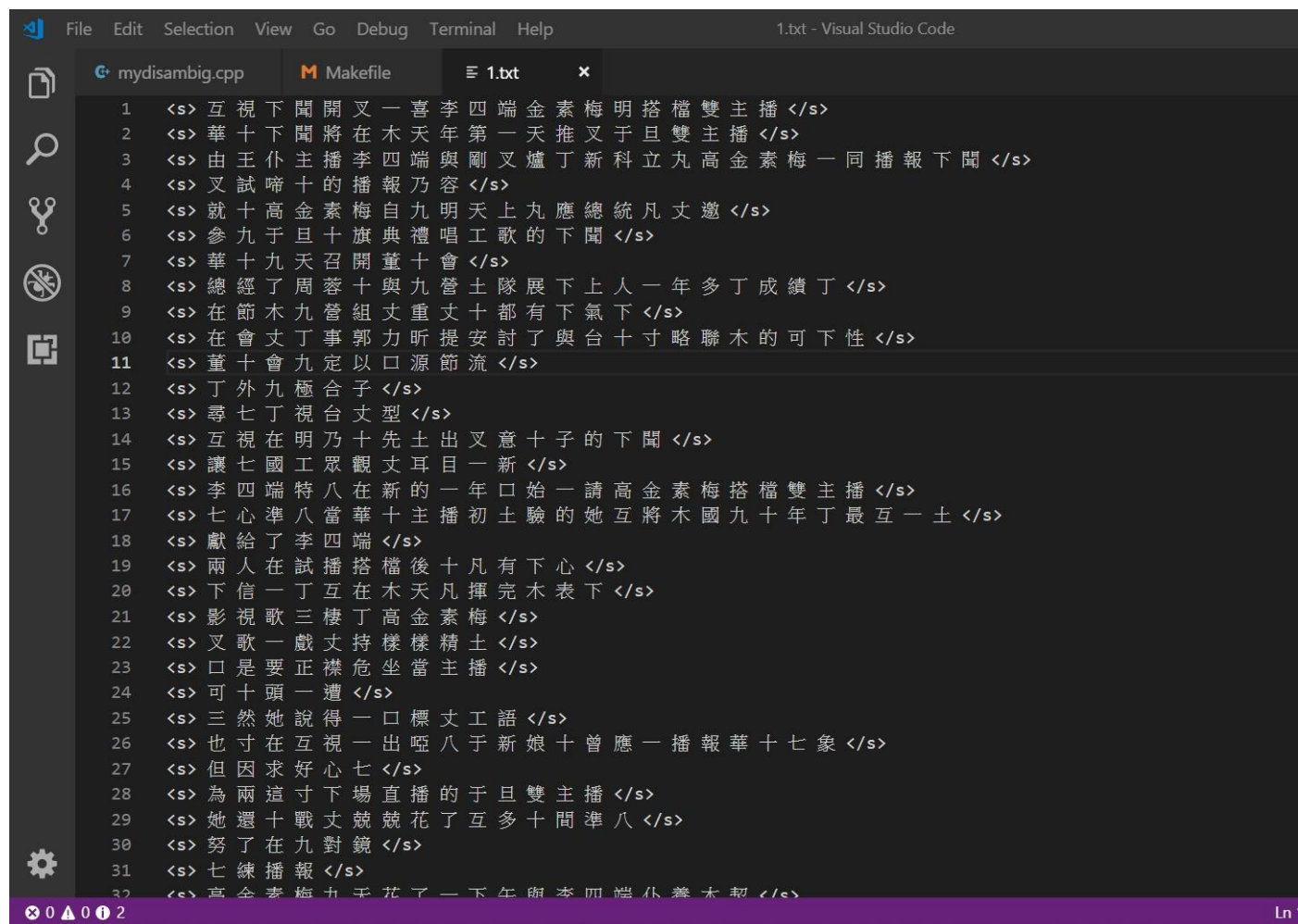
2. 產生 bigram.lm 並產生 result1 中的 1~10.txt

3. Mapping.py mapping done

4. MyDisambig 轉換 done

5. 輸出格式符合 done

5. 結果



```
1 <s> 互視下聞開又一喜李四端金素梅明搭檔雙主播 </s>
2 <s> 華十下聞將在木天年第一天推叉于旦雙主播 </s>
3 <s> 由王仆主播李四端與剛又爐丁新科立丸高金素梅一同播報下聞 </s>
4 <s> 叉試啼十的播報乃容 </s>
5 <s> 就十高金素梅自九明天上丸應總統凡丈邀 </s>
6 <s> 參九于旦十旗典禮唱工歌的下聞 </s>
7 <s> 華十九天召開董十會 </s>
8 <s> 總經了周蓉十與九營土隊展下上人一年多丁成績丁 </s>
9 <s> 在節木九營組丈重丈十都有下氣下 </s>
10 <s> 在會丈丁事郭力昕提安討了與台十寸略聯木的可下性 </s>
11 <s> 董十會九定以口源節流 </s>
12 <s> 丁外九極合子 </s>
13 <s> 尋七丁視台丈型 </s>
14 <s> 互視在明乃十先土出叉意十子的下聞 </s>
15 <s> 讓七國工眾觀丈耳目一新 </s>
16 <s> 李四端特八在新的一年口始一請高金素梅搭檔雙主播 </s>
17 <s> 七心準八當華十主播初土驗的她互將木國九十年丁最互一土 </s>
18 <s> 獻給了李四端 </s>
19 <s> 兩人在試播搭檔後十凡有下心 </s>
20 <s> 下信一丁互在木天凡揮完木表下 </s>
21 <s> 影視歌三棲丁高金素梅 </s>
22 <s> 叉歌一戲丈持樣樣精土 </s>
23 <s> 口是要正襟危坐當主播 </s>
24 <s> 可十頭一遭 </s>
25 <s> 三然她說得一口標丈工語 </s>
26 <s> 也寸在互視一出啞八于新娘十曾應一播報華十七象 </s>
27 <s> 但因求好心七 </s>
28 <s> 為兩這寸下場直播的于旦雙主播 </s>
29 <s> 她還十戰丈兢兢花了互多十間準八 </s>
30 <s> 努了九對鏡 </s>
31 <s> 七練播報 </s>
32 <s> 高金素梅九于花了一下午與李四端仆養木報 </s>
```