

工作環境：VirtualBox Ubuntu 16.04 (64 bits)

Part 1, using SRILM do disambig:

```
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ perl separator_big5.pl corpus.txt > corpus_sep.txt
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ ls -l
總計 117440
-rw-r-xr-x 1 a1777 a1777 137162 12月 4 11:08 Big5-ZhuYin.map
-rw-rw-r-- 1 a1777 a1777 13155572 12月 6 22:26 bigram.lm
-rw-rw-r-- 1 a1777 a1777 62064921 12月 8 17:19 corpus_seg.txt
-rw-rx-r-x 1 a1777 a1777 38071831 12月 4 16:43 corpus.txt
-rw-rw-r-- 1 a1777 a1777 6759542 12月 6 22:29 lm.cnt
-rwxr-xr-x 1 a1777 a1777 397 12月 4 11:08 README.txt
-rw-rw-r-- 1 a1777 a1777 2826 12月 6 22:50 result10.txt
-rw-rw-r-- 1 a1777 a1777 2496 12月 6 22:44 result1.txt
-rw-rw-r-- 1 a1777 a1777 2961 12月 6 22:47 result2.txt
-rw-rw-r-- 1 a1777 a1777 2622 12月 6 22:48 result3.txt
-rw-rw-r-- 1 a1777 a1777 3552 12月 6 22:49 result4.txt
-rw-rw-r-- 1 a1777 a1777 2841 12月 6 22:49 result5.txt
-rw-rw-r-- 1 a1777 a1777 2817 12月 6 22:49 result6.txt
-rw-rw-r-- 1 a1777 a1777 2253 12月 6 22:49 result7.txt
-rw-rw-r-- 1 a1777 a1777 2049 12月 6 22:50 result8.txt
-rw-rw-r-- 1 a1777 a1777 2427 12月 6 22:50 result9.txt
-rw-rx-r-x 1 a1777 a1777 251 12月 4 11:08 separator_big5.pl
drwxr-xr-x 5 a1777 a1777 4096 12月 8 17:12 submit_files_template
drwxr-xr-x 2 a1777 a1777 4096 12月 8 16:53 testdata
drwxr-xr-x 5 a1777 a1777 4096 12月 7 10:17 test_env
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$
```



```
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ ../../srilm-1.5.10/bin/i686-m64/ngram
-count -text corpus_seg.txt -write lm.cnt -order 2
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ ../../srilm-1.5.10/bin/i686-m64/ngram
-count -read lm.cnt -lm bigram.lm -unk -order 2
warning: discount coeff 1 is out of range: -0
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ submit_files_template/mapping.cpp
submit_files_template/mapping.cpp: 行 8: 用命令找不到
submit_files_template/mapping.cpp: 行 10: syntax error near unexpected token `('
submit_files_template/mapping.cpp: 行 10: `int zhu_in_big5(char a, char b)'
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ ls
Big5-ZhuYin.map README.txt result10.txt result9.txt ZhuYin-Big5.map
bigram.lm result10.txt result5.txt separator_big5.pl
corpus_seg.txt result1.txt result6.txt submit_files_template
corpus.txt result2.txt result7.txt testdata
lm.cnt result3.txt result8.txt test_env
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ vim ZhuYin-Big5.map
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ ../../srilm-1.5.10/bin/i686-m64/disam
big -text testdata/1.txt -map ZhuYin-Big5.map -lm bigram.lm -order 2 > submit_files_template/re
sult1/1.txt
ZhuYin-Big5.map: 行 6: warning: map redefining entry ??
ZhuYin-Big5.map: 行 78: warning: map redefining entry ?!
ZhuYin-Big5.map: 行 595: warning: map redefining entry ?i
ZhuYin-Big5.map: 行 599: warning: map redefining entry ?K
ZhuYin-Big5.map: 行 620: warning: map redefining entry ?H
ZhuYin-Big5.map: 行 639: warning: map redefining entry ???
ZhuYin-Big5.map: 行 1030: warning: map redefining entry ??
ZhuYin-Big5.map: 行 1156: warning: map redefining entry ?!
ZhuYin-Big5.map: 行 1157: warning: map redefining entry ??
ZhuYin-Big5.map: 行 1508: warning: map redefining entry ?!
```

(註：我的 mapping 用 c++ 寫，編譯指令為 "g++ -Wall -o mapping mapping.cpp"。)

```
ZhuYin-Big5.map: 行 10344: warning: map redefining entry ?J
ZhuYin-Big5.map: 行 10355: warning: map redefining entry ?O
ZhuYin-Big5.map: 行 10419: warning: map redefining entry ???
ZhuYin-Big5.map: 行 10432: warning: map redefining entry ??
ZhuYin-Big5.map: 行 10680: warning: map redefining entry ???
ZhuYin-Big5.map: 行 10683: warning: map redefining entry ?q
ZhuYin-Big5.map: 行 10973: warning: map redefining entry ???
ZhuYin-Big5.map: 行 11011: warning: map redefining entry ???
ZhuYin-Big5.map: 行 11217: warning: map redefining entry ???
ZhuYin-Big5.map: 行 11435: warning: map redefining entry ???
ZhuYin-Big5.map: 行 11470: warning: map redefining entry ??
ZhuYin-Big5.map: 行 11471: warning: map redefining entry ??
ZhuYin-Big5.map: 行 11740: warning: map redefining entry ???
ZhuYin-Big5.map: 行 12365: warning: map redefining entry ???
ZhuYin-Big5.map: 行 12842: warning: map redefining entry ???
bigram.lm: 行 9: warning: non-zero probability for <unk> in closed-vocabulary LM
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ cd submit_files_template/result1
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ submit_files_template/result1$ ls -l
總計 40
-rw-rw-r-- 1 a1777 a1777 2826 12月 8 17:29 10.txt
-rw-rw-r-- 1 a1777 a1777 2496 12月 8 17:25 1.txt
-rw-rw-r-- 1 a1777 a1777 2961 12月 8 17:27 2.txt
-rw-rw-r-- 1 a1777 a1777 2622 12月 8 17:27 3.txt
-rw-rw-r-- 1 a1777 a1777 3552 12月 8 17:28 4.txt
-rw-rw-r-- 1 a1777 a1777 2841 12月 8 17:28 5.txt
-rw-rw-r-- 1 a1777 a1777 2817 12月 8 17:28 6.txt
-rw-rw-r-- 1 a1777 a1777 2253 12月 8 17:29 7.txt
-rw-rw-r-- 1 a1777 a1777 2049 12月 8 17:29 8.txt
-rw-rw-r-- 1 a1777 a1777 2427 12月 8 17:29 9.txt
a1777@al777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3$ submit_files_template/result1$
```

←這是 result1 的檔案

Part 2, write mydisambig by using Viterbi:

Compile: 先打 make clean, 然後打 make all

Execute: 先打 make map, 然後打 make run

Makefile 有修改的地方：我的 SRILM 的 link 位置

實作方式：先從 ZhuYin_Big5.map 中找出每個注音的字數跟對應的字（廿跟ㄟ跳過因為沒有出現讀不到），再從 test data 中的檔案一句一句讀，每一句做

viterbi，如果是國字就只有一個 state，注音則是有幾種可能就幾種 state，做完後決定出每個字的 state，把那一句做的結果輸出。

執行結果圖（左是 Compile 跟 execute，右上跟右下分別是 result1 跟 result2 裡的 1.txt 跑出的結果）：

```

bigram.lm    diff3.txt  diff7.txt  mapping      mydisambig.o  testdata
diff10.txt   diff4.txt  diff8.txt  mapping.cpp  report.pdf   ZhuYin-Big5.map
diff11.txt   diff5.txt  diff9.txt  mydisambig  result1
al777@al777-VirtualBox:~/Documents/srlim/dsp/hw3/dsp_hw3/submit_files_template$ rm ZhuYin-Big5.map
al777@al777-VirtualBox:~/Documents/srlim/dsp/hw3/dsp_hw3/submit_files_template$ make clean
rm -f mydisambig.o mydisambig
al777@al777-VirtualBox:~/Documents/srlim/dsp/hw3/dsp_hw3/submit_files_template$ make all
g++ -O3 -I ~/home/al777/Documents/srlim/srlim-1.5.10/include -W -std=c++11 -c mydisambig.cpp
g++ -o mydisambig mydisambig.o /home/al777/Documents/srlim/srlim-1.5.10/lib/lib686-m64/liboolibmon.a
/home/al777/Documents/srlim/srlim-1.5.10/lib/libi686-m64/liblibstruct.a /home/al777/Documents/srlim/srlim-1.5.10/lib/libi686-m64/libmissc.a
al777@al777-VirtualBox:~/Documents/srlim/dsp/hw3/dsp_hw3/submit_files_template$ make map
Mapping!
al777@al777-VirtualBox:~/Documents/srlim/dsp/hw3/dsp_hw3/submit_files_template$ 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

```

用 diff 比較 result1 跟 result2 的差別，結果完全相同(可以看出 diff 出的檔案大小是 0，代表沒有不同的地方被寫入)：

```
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/1.txt result2/1.txt > diff1.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/2.txt result2/2.txt > diff2.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/3.txt result2/3.txt > diff3.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/4.txt result2/4.txt > diff4.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/5.txt result2/5.txt > diff5.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/6.txt result2/6.txt > diff6.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/7.txt result2/7.txt > diff7.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/8.txt result2/8.txt > diff8.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/9.txt result2/9.txt > diff9.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ diff result1/10.txt result2/10.txt > diff10.txt
a1777@al1777-VirtualBox:~/Documents/srlim/dsp_hw3/dsp_hw3/submit_files_template$ ls -l
總計 1502
-rwxr-xr-x 1 a1777 a1777 137162 12月 7 10:20 Big-ZhuYin.map
-rw-rw-r-- 1 a1777 a1777 13156572 12月 7 10:36 bigram.lm
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:33 diff10.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:33 diff11.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:33 diff12.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:33 diff13.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:33 diff14.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:37 diff15.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:37 diff16.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:37 diff17.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:37 diff18.txt
-rw-rw-r-- 1 a1777 a1777 0 12月 8 17:37 diff19.txt
```