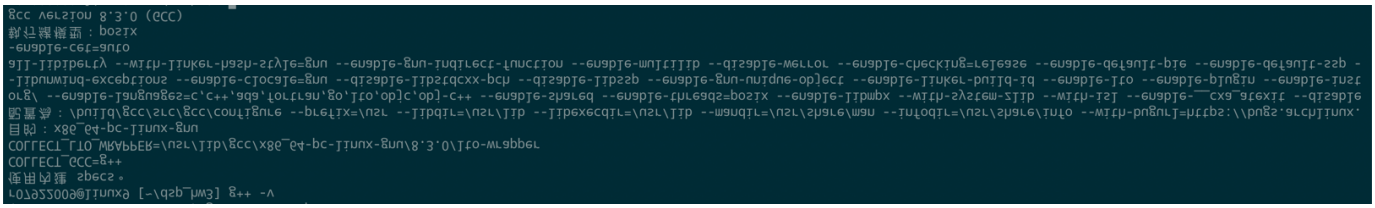
數位語音處理概論作業三

資工所碩一 R07922009 丁國騰

一、環境：

本次作業是在 CSIE workstation 完成，使用 linux9.csie.ntu.edu.tw 工作站（gcc 版本為 8.2.1）



二、How to “compile” your program

將 Makefile 中的 SRIPATH 與 MACHINE_TYPE 設定好，使用 make 指令（即「輸入 make 於終端介面」）即可，或也可以在設定好環境變數的情況下使用以下指令執行：

```
g++ -O3 -I$(SRIPATH) -w --std=c++11 -c mydisambig.cpp
g++ -o mydisambig mydisambig.o $(SRIPATH)/lib/$(MACHINE_TYPE)/liboolm.a
$(SRIPATH)/lib/i$(MACHINE_TYPE)/libdstruct.a /$(SRIPATH)/lib/i$(MACHINE_TYPE)/libmisc.a
```

二、How to “execute” your program

1. make map: 執行 mapping.py，產生 ZhuYin-Big5.map

   ```
   python mapping.py Big5-ZhuYin.map ZhuYin-Big5.map
   ```

2. make build_lm: 使用 srilm 提供的 ngram-count 對已分割過的檔案製造出 bigram.lm

   ```
   perl separator_big5.pl corpus.txt > corpus_seg.txt
   $(SRIPATH)/bin/$(MACHINE_TYPE)/ngram-count -text corpus_seg.txt -write lm.cnt -order 2
   $(SRIPATH)/bin/$(MACHINE_TYPE)/ngram-count -read lm.cnt -lm bigram.lm -unk -order 2
   ```

   如果沒有對 txt 進行分割請先執行 mv testdata testdata_old、perl separator_big5.pl testdata_old/${i}.txt > testdata/${i}.txt

3. make test: 使用 srilm 提供的 disambig 解析出答案，並存於$(TESTDIR)，預設為 result1

   ```
   [ -d $(TESTDIR) ] || mkdir -p $(TESTDIR);
   @for i in $(shell seq 1 10); do \
       $(SRIPATH)/bin/$(MACHINE_TYPE)/disambig -text $(TESTDATA)/$${i}.txt -map $(TO) -lm
   $(LM) -order 2 > $(TESTDIR)/$${i}.txt; \
   done;
   ```

4. make run: 執行 mydisambig 並把結果存於$(MYDISAMBIGDIR)，預設為 result2

   ```
   [ -d $(MYDISAMBIGDIR) ] || mkdir -p $(MYDISAMBIGDIR);
   for i in $(shell seq 1 10) ; do \
       echo "Running $$i.txt"; \
       ./mydisambig -text testdata/$$i.txt -map $(TO) -lm $(LM) -order 2 >
   $(MYDISAMBIGDIR)/$$i.txt; \
   done;
   ```

5. make clean: 清除之前 make 產生的檔案

   ```
   rm -f mydisambig.o mydisambig
   ```

三、What you have done

我按照課程網站的 Q&A 修正了 makefile，並且以 SRILM 將注音轉為國字

並在 mydisambig 中實作了 viterbi 演算法，並對其結果與 SRILM 的結果比對