Digital Speech Processing - Homework #3

資工三 B05902118 陳盈如

- Enviornment

CSIE workstation

- How to "compile"

- ***Make map*** *can generate ZhuYin-Big5.map from the given Big5-ZhuYin.map by the command, "python3 mapping.py $(FROM) $(TO)".*
- ***Make*** *can compile mydisambig.cpp.*

- How to "execute"

- ***Make run*** *can make a directory by the command, "mkdir result2", and run my code toward different txt by the command, "./mydisambig -text testdata/$$i.txt -map $(TO) -lm $(LM) -order 2 >result2/$$i.txt".*
- ***Make clean*** *can delete "mydisambig.o" and the execution file, "mydisambig".*

- What I have done

mapping.py:

1. *Read in the Big5-ZhuYin.map.*
2. *Get the character and its ZhuYin.*
3. *Put the character into dictionary depending on its ZhuYin. If the ZhuYin has appeared, just append the character. If not, create this ZhuYin and record the character.*

4. *Sort the dictionary in "ㄅㄆㄇㄈ……" order.*

5. *Write ㄅㄆㄇ…… and characters into ZhuYin-Big5.map.*

mydisambig.cpp:

1. *Take advantage of structure.*
   (1) *Make a line with many word vectors.*
   (2) *Each word vector records its own character and all possible characters behind it.*
   (3) *Each character is represented by char[ ].*
   (4) *Use the similar structure to record the possibility.*
2. *Use the function given on the course website to call language model*
3. *Read in the TestData and get one line in the file each time.*
4. *Utilize the structure built in advance to record each word in a line and the possible words for ZhuYins in this line.*
   (ps. We can get the possible words from ZhuYin-Big5.map.)
5. *Implement Viterbi algorithm and record the index of the highest possibility simultaneously.*
6. *Go through the highest possibility path and get the characters which we need.*
   (ps. Don't forget to add <s> at the beginning and </s> at the end of each line.)