

# HW3 Report

b04501127 凌于凱

## Enviroment:

os: ubuntu 16.04

g++: gcc version 5.4.0

## How to “compile” my program:

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

## How to “execute” your program:

1. copy you Big5-ZhuYin.map to the directory

```
$ cp $(your Big5-ZhuYin.map) ./Big5-ZhuYin.map
```

2. make ZhuYin-Big5.map file

```
$ make map
```

3. execute ./mydisambig

```
./mydisambig -text $(testfile) -map $(Big5-ZhuYin mapping file) -lm $(bigram model) -order 2 > $(outputfile)
```

## What I have done:

1. mapping.py

透過套件 re 將字串分割，再透過 dict()將每個字存入對應的注音或是字的字典裡

2. mydisambig.cpp

(1). 利用 Srilm 內建函式讀取 bigram model

(2). 利用 std map 建立注音-文字的字典

(3). 將句子分割成一個個詞，可透過 Vocab::parseWords 函式

(4). 透過 Viterbi 演算法找出每一條路徑對應的機率

(5). 透過 back tracking 找到機率最高對應的句子。