DSP_hw3

b06902111 資工三 林慶珠

1. run my program

- generate language model

  in terminal, the following command will output `lanmodel.lm` as language model

  ```
  $ make lanmodel
  ```

- generate 注音 to 字 mapping

  `$1` is the Big5-ZhuYin mapping file and `$2` is the output ZhuYin-Big5 mapping file

  ```
  $ make map FROM=$1 TO=$2
  ```

- decode 注音文

  `$1` is the segemented file to be decoded, `$2` is the ZhuYin-Big5 mapping file, `$3` language model file and `$4` is the decoded output file

  ```
  ./mydisambig $1 $2 $3 $4
  ```

- I `diff` the decoded output file with `mydisambig` and SRILM `disambig` (with perled testdata as input in both commands) , it turns out their outputs are exactly the same.

2. running time

I use `time` command to calculate execution time as follows.

```
(srilm) root@60e5afb4b456:/opt/dsp_hw3# time ./mydisambig test_data/3_perled.txt ZhuYin-Big5.map lanmodel.lm test_
data/3_out.txt
lanmodel.lm: line 9: warning: non-zero probability for <unk> in closed-vocabulary LM

real    0m13.370s
user    0m13.340s
sys     0m0.020s
(srilm) root@60e5afb4b456:/opt/dsp_hw3#
```

file	execution time
1.txt	13.602 s
2.txt	22.182 s
3.txt	13.370 s
4.txt	18.561 s
5.txt	13.823 s
6.txt	27.593 s
7.txt	20.392 s
8.txt	22.016 s
9.txt	24.488 s
10.txt	37.722 s
example.txt	19.922 s
<avg>	$\sim$ 21.242 s

3. Reference

- b06902017
- Internet: c++, python, Makefile