DSP HW3

ZhuYin-mixed sequence decoding

B04902090 資工三 施長元

Environment & Execution:

1. Your environment (CSIE workstation, Cygwin, …)

   Windows Linux Subsystem(WSL) Ubuntu 16.04 on Windows 10

   g++ version: 5.4.0 20160609 (Ubuntu 5.4.0-6ubuntu1~16.04.9)

   Ryzen 5 1600, DDR4 2400MHz, SSD, GPU doesn't matter.

2. How to "compile" your program

   `make MACHINE_TYPE=[MACHINE_TYPE] SRIPATH=[SRIPATH] all`

3. How to "execute" your program
   - **`make map`**

     (Generate `ZhuYin-Big5.map` from my python script)
   - **`make MACHINE_TYPE=[MACHINE_TYPE] SRIPATH=[SRIPATH] run`**

     (Run `mydisambig` and output the result in `result2/`)

Report:

4. What you have done
   - Use `separator_big5.pl` to separate character for `corpus`、`testdata/[1..10].txt`
   - `make map` to Generate ZhuYin-Big5.map
     - A python script
     - Store `ZhuYin-Big5` pairs in dictionary for avoiding duplication
     - big5hkscs works better than big5
   - Use provided binary SRILM (Debian x86_64 executable files) to run result1
     - `lm.cnt`、`bigram.lm`、`result1/`
   - And I realized that I need to compile SRILM to read the source codes/headers
     - Compile the code
     - An error encountered: `sudo apt install csh`
   - `Mydisambig.c`:
     - API from `SRILM/include/`:
       - `File.h`
         - `maxWordsPerLine`
       - `Prob.h`
         - `LogP`、`Prob`
         - `ProbToLogP`

        - LogP_Zero
    - Ngram.h
        - Ngram for language model
        - Ngram.read
    - Vocab.h
        - Vocab.getIndex
        - VocabString
        - VocabIndex
        - Vocab_None
        - Vocab::parseWords
    - VocabMap.h
        - Map declaration
        - VocabMap.read
        - VocabMap.getWord
        - VocabMapIter
        - unkIndex
    - Trellis.cc
        - Vector liked structure storing things
        - TrellisIter
        - trellis.step
        - trellis.update
        - trellis.setProb
        - trellis.viterbi
- Then it's just a matter of time to put these tools together.
- Trellis.cc plays a big role in this assignment, Finish Viterbi for just a row.