

DSP_hw3 report

朱世耘 r09922115

- mapping.py:

- 1.使用類似hashing的方法，把國字放入對應的注音中。
- 2.把每個注音對應到的字印出來。
- 3.讓每個國字對應到自己，再印出來。（這邊如果有破音字，會有重複定義的情況發生）

mydisambig.cpp:

- 1.處理argv，一行一行讀檔。
 - 2.把每行文字丟入viterbi algorithm。
- Viterbi algorithm:
- 3.把每個注音可能對應到的國字存起來。
 - 4.計算相鄰國字出現的機率，把機率相乘用dynamic programming計算local最佳路徑。
 - 5.一邊計算一邊把過程存在backtrace用的陣列。
 - 6.到底了之後，找出global的最佳路徑，back trace回去。
 - 7.印出整個句子

result:

以1.txt為例：

```
<s> 忽視新聞開場迎喜李四端金素梅明搭檔雙主 播 </s>
<s> 華社新聞將將在明四天年第一天推出梅約旦雙主 播 </s>
<s> 由王牌主播李四端與剛出爐的新科立委高金素梅也同播報新聞 </s>
<s> 嘗試是加元金旦素梅自典禮唱天上歌的應總統府之邀 </s>
<s> 參加社今天召開董事會 </s>
<s> 總經理周蓉參與經營團隊現上任一年多達成績單 </s>
<s> 在節目經營組織重鎮市都有新氣象 </s>
<s> 在會長董事郭力昕提案討論與台商策略聯盟的可行性 </s>
<s> 董事會決定以客源節流 </s>
<s> 對外積極合作 </s>
<s> 尋求電視台中型 </s>
<s> 忽視在明年初先提出誠意十字的新聞 </s>
<s> 讓全國觀眾觀眾耳目一新 </s>
<s> 李四端特別在華社主 播 初體 驗 的 她 還 將 美 國 九 十 年 度 最 後 一 天 </s>
<s> 全 心 準 備 當 四 端 </s>
<s> 兩 人 在 一 試 播 檔 後 十 分 有 信 心 </s>
<s> 相 信 定 會 在 明 天 發 揮 完 美 表 現 </s>
<s> 影 視 三 樓 的 高 金 素 梅 </s>
<s> 唱 歌 演 戲 主 持 樣 樣 精 通 </s>
<s> 可 是 要 正 襟 危 坐 當 主 播 </s>
```

可以觀察到還是會有錯字，像是：忽視（應為華視）、中型（重視），但是都是日常中出現機率高的詞，或許bigram有他的極限在，如果加入trigram或其他輔助，可以表現得比較好。