

DSP HW3 report

b07902048 資工三 李宥霆

在這次作業中做了什麼

我在這次的作業中實做了bigram的部份，基本上使用mydisambig與disambig去跑所有的test data，出來的結果都是一模一樣的，執行時間上來說也很接近，顯示實做出來的程式應該跟disambig是差不多的。

實做中遇到的問題

1

在實做過程中遇到最大的問題應該就是中文詞語的讀寫，有發現一個問題，是當我們把部份單詞使用string的型態接住，在將string放入structure，之後再從structure中把字元拿出來時，有少數的字元會變成亂碼，不確定是否是因為實做的過程中有誤或是存在某些潛在的問題，之後改成不用structure後就恢復正常了。

2

另外，在big5-zhuYin.map這個檔案中也有出現一個奇怪的編碼問題，就明明"響"這個字元有出現在big5-zhuYin.map之中，但再把檔案內容讀進C++中之後會顯示找不到"響"這個字元，猜測可能是有某些字元的編碼方式有兩種以上，雖然顯示出來的是同一個字，但其實編碼是不同的，導致有些明明存在的字元被判定為<unk>。