**DSP HW3**

**資工所 R05922130 王瀚磊**

1. **Introduction**

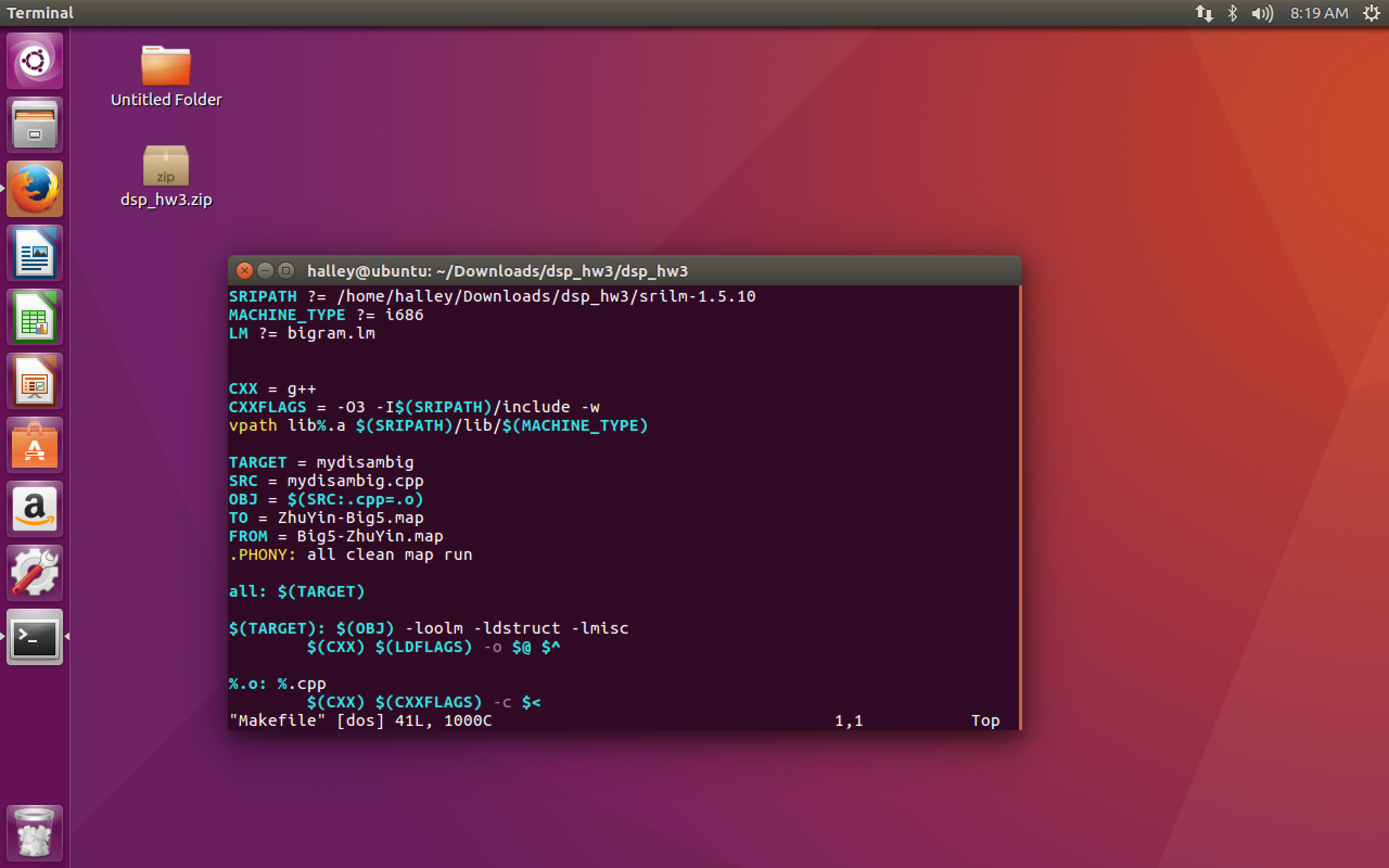
本次作業為實作辨識注音文，從（注音/國字）轉換成（國字），其中利用到TA所提供的toolkit：SRILM-1.5.10。在產生mapping的dictionary後可以利用Viterbi找出最佳的path並且判斷注音所對應到的國字。

1. **Program Environment**

Mac OSX / Ubuntu i686 LTE 14 (VMWare Workstation)

1. **How To Run?**

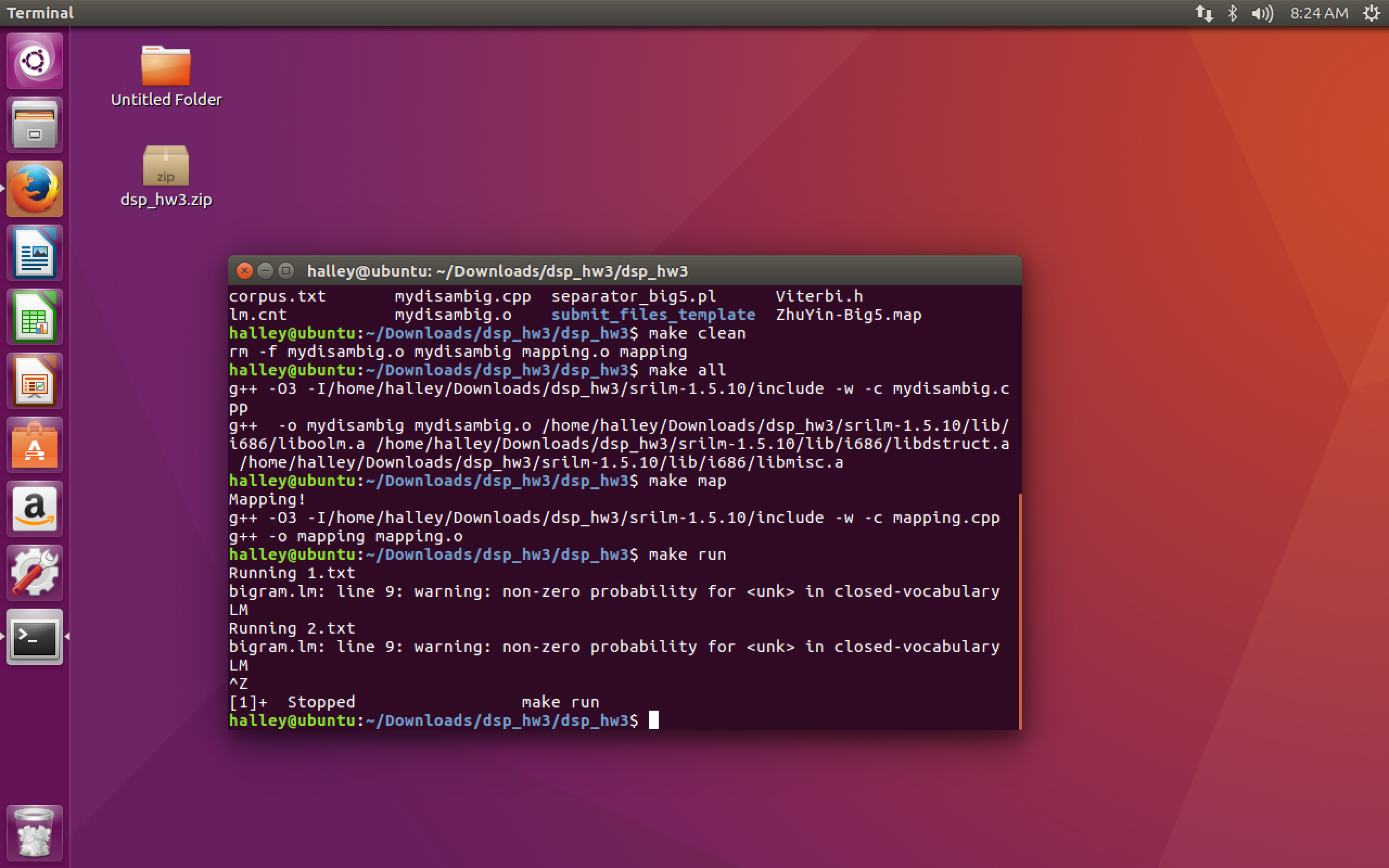
首先更改Makefile中的SRIPATH，將其路徑改為SRILM所在的資料夾位置，如圖所示。



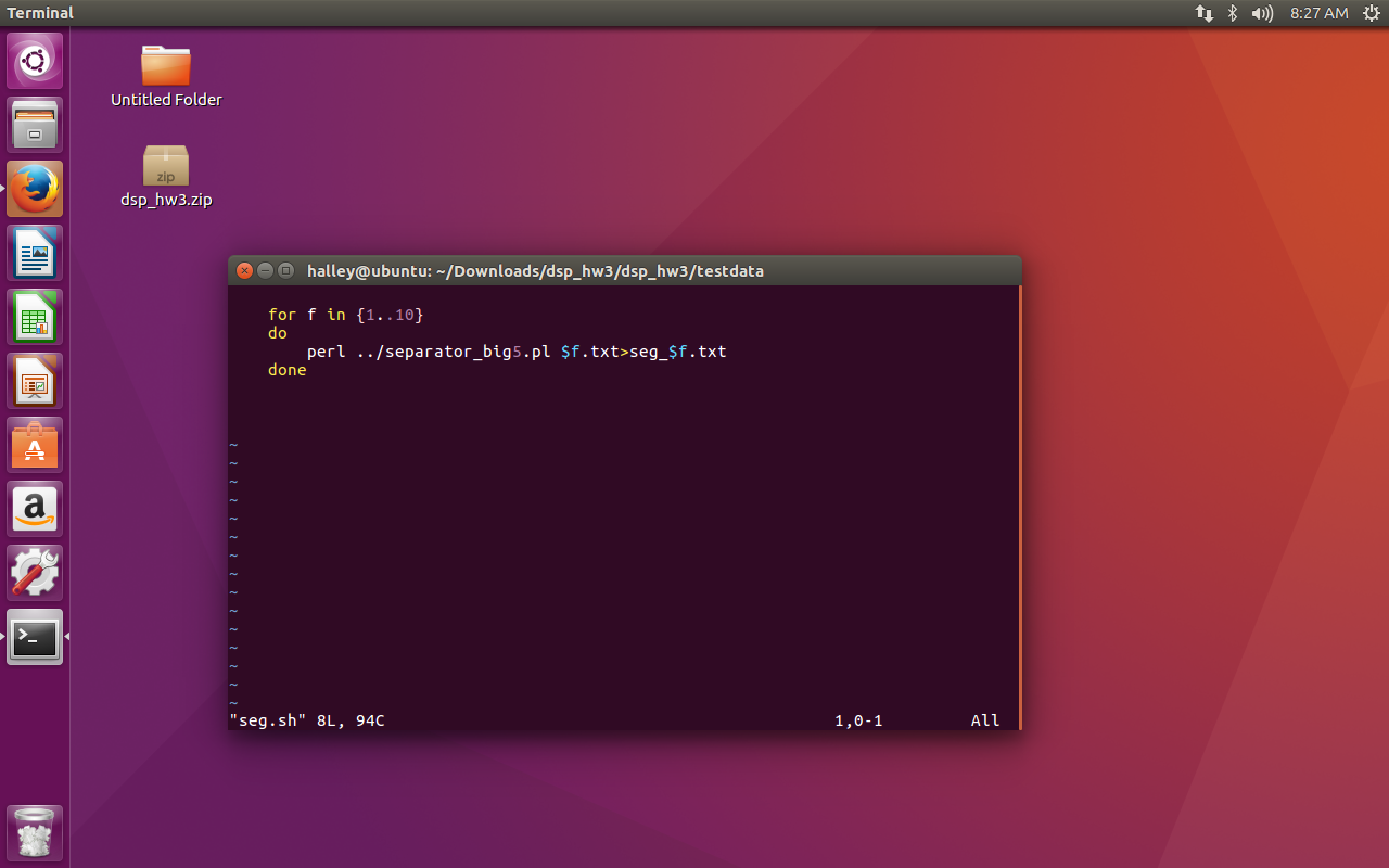
接著即可輸入makefile 指令，如下所示：

* make clean
* make all，在此已經預設MACHINE\_TYPE=i686-m64 SRIPATH=/home/ta/srilm-1.5.10
* make map
* make run，在此已經預設MACHINE\_TYPE=i686-m64 SRIPATH=/home/ta/srilm-1.5.10 LM=bigram.lm

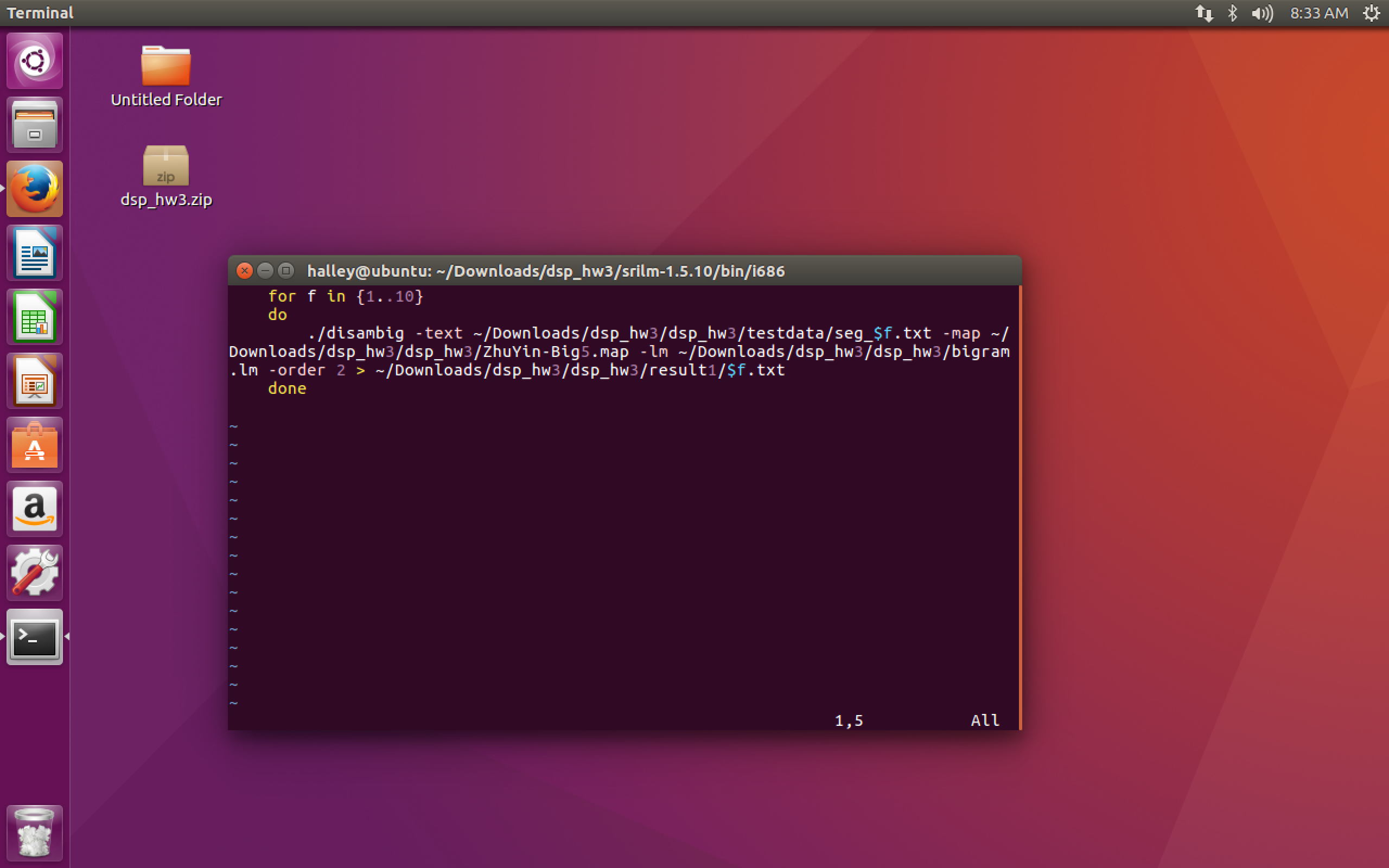
結果如下圖所示。



1. **Step by Step**
2. 首先先安裝srilm，基本上順著網頁上的QA即可完成，然而要注意在修改stdio.h中的\_\_off64\_t取代成long int時，要注意不要修改到其他相似關鍵字如\_\_off64\_t\_defined，否則會出現錯誤，此外我還需要額外安裝g++才能順利安裝完成。
3. 接著安裝zhTW.BIG5，並且修改vimrc新增其encoding類型:big5，如此才能用vim順利顯示中文檔案。
4. 接著將助教給的test data做segmentation，將每個字以空白間隔，在這邊我自己寫一個script將十筆data做segmentation，如圖所示。



1. 接下來即可利用安裝後的檔案（disambig）來執行判斷結果，這邊一樣寫一個script讓他執行後將結果輸出至result1資料夾底下，如圖所示。



1. 接著撰寫mapping.cpp，將Big5-Zhuyin 轉換至 Zhuyin-Big5 ，基本上就是讀取注音的第一個音，利用vector將相同的音所對應的字詞存取。
2. 最後撰寫mydisambig.cpp，首先先寫一個viterbi計算機率最大的那條path，並且利用剛剛安裝好的檔案Ngram.h以及轉置好的mapping檔案 Zhuyin-Big5.map以及 利用ngram-count得到的bigram.lm即可得到所有的language model並且算出max probability path。