




Programming Assignment 1

- Write a program to extract terms from a document.
 - You have to do:
 - Tokenization. 
 - Lowercasing everything.
 - Stopword removal. 
 - Stemming using Porter's algorithm. 
 - Save the result as a txt file.
 - Text collection:
 - One English news documents
(<https://ceiba.ntu.edu.tw/course/b079e8/content/1.txt>).
 - **Please zip and submit ¹the result of document 1, ²the source code, and ³a report to TA.**
 - 2 weeks to complete, that is **2012/10/2**.