

# IR Term Project Proposal

## -- 文E青年(藏頭文產生器)

B00705010 資管四 劉力源

B00705024 資管四 周詮儒

B00705045 資管四 張奕儒

B00705047 資管四 施淮振

### 前言

藏頭文是近年來在網路上火紅的惡搞文章格式，在看似沒有其他意思的文章，按著特別的方式念，可以讀出隱藏的訊息，鄉民經常使用藏頭文來批評別人。藏頭文有各式各樣的排列方式，藏的方式通常都是每行第一個字直的念下來，或者是斜排的念，還有厲害的網民曾經用過M的排列方式來藏頭。

我們這組做的主題就是藏頭文產生器，在從論壇搜集了大量文章存進IR System之後，使用者輸入想要隱藏的訊息，即可產生出藏頭文。除此之外還可以做出其他的應用，例如藏頭文猜謎，考驗使用者的眼力來找出隨機產生文章裡的藏頭文。

以下為藏頭文的範例：

拜官求侯皆父意，託鄰走里六百地，不樂綿雨絲絲細，要舒喜，讓道歸府另擇期。  
勝機迷茫空仇嫉，文誅筆伐聲聲厲，不看己身滿瘡痍。開智囊，心頭滿是怨毒計。

出處：#1KOkvRsE (Gossiping) -- PTT實業坊

### possible solution

我們的主要去建一個詞庫跟句庫，資料來源可以從網路上論壇(如：PTT實業坊，大量下載文章，並解析產生詞庫跟句庫所需要的資料。

產生藏頭文時，先把藏頭文的資訊做拆解，再根據每一個字，從句庫裡面找尋相對應的句子並組合：若在搜尋句庫時沒有找到沒有適當的句子，則從詞庫中，取出一些詞，再用這個詞的詞性跟CFG(Context free grammar)去產生一個句子，若某些字極為稀有，難以從資料庫中找尋對應的句子或組合成句子，則會將該字用諧音的方式代換成另一個字，再重啟組成句子的步驟。確認藏頭資訊在文章無誤後即可輸出，一篇藏頭文就誕生啦。