

關於前處理，前處理應該是最大的問題。後面 LSTM 我就是套網路上面的找到的 **model** 而已。

- (1) 建立屬於這次問題的辭庫，這是屬於前處理範圍，所以 **testing data** 沒有 **label** 也沒關係。主要是把 **key word** 裡面的辭彙都輸入 **dict** 裡面。
- (2) 再接下來就是把每筆資料轉換成一個一個詞(包含 **title** 跟 **key word**)
- (3) 之後轉成 **vector**，但是因為每一筆資料長度不太一樣，所以我們要用 **padding** 把每筆資料 **x** 變得一樣長。這邊設定 30
- (4) 再來就是 **LSTM** 模型，然後預測，交卷。