

作業心得：

其實上課聽老師講理論時我幾乎都聽不懂，只大概知道是將單詞轉成向量並藉由判斷兩個向量的關係決定兩個詞的相似度，我還真沒想到自然語言居然和向量扯得上關係，挺酷的，不過其他太多的數學模型就讓我有點不知所措。要實作時，我開始在網路上不停地搜尋資料，希望可以先找到一些 **gensim** 的建模方法，建完模之後就可以根據模型尋找相似字，著實滿有趣的。作業要求要根據給定的句子繼續延伸，利用哈利波特小說為素材來編寫故事，我建完 **word2vec** 的模型後要開始從前後文來計算機率，也就是某一字出現另外兩字中間的機率，尋找一個最高的填進所有編寫的故事裡，持續不斷。我一直沒有辦法處理好標點符號的問題，甚至在建完模型還會跑出莫名的 '+' 號，讓我百思不得其解，我認為應該我是在 **tokenize** 單詞時有出錯。總之最後我有印出一些單詞，雖然標點符號有點亂髮無章，有些句型也不太對，不過依稀可以看到有幾句符合文法的句子。總體來說我算是沒有做出來吧，不過第一次了解自然語言原來是用向量處理的觀念覺得挺特別的，經過這次多少更了解了自然語言處理，也是有學到一些東西。