**6/2人工智慧 S10767007林煒翔**

應用：意見探勘

文本分析：對新聞、科技論文、電子郵件、網頁、部落格貼文、微網誌、病歷等不同類型來源的數位資料，進行文件探勘

例如：市場產品資訊、政治意見追蹤、社群網路分析、熱門議題分析。

自然語言處理的一些典型問題：

1. 歧異解析：詞彙層次、語法層次、語意層次
2. 容錯力：a.打錯字b.語法錯誤c.未知詞（沒有被收錄詞彙）
3. 強健性：a.領域改變b.網路語言：方言俗語、各門外語、縮略語、諧音、符號合併
4. 表情符號

自然語言處理一些任務

1. 詞素—詞素分析
2. 詞彙—中文斷詞
3. 命名實體—命名實體辨析
4. 詞性—詞性標記
5. 句法類別—句法剖析
6. 相依類別—相依剖析
7. 語意類別—語意關係標記
8. 言談關係—言談剖析
9. 意見—意見探勘
10. 情感—情感分析
11. 立場—立場偵測

在向量計算的部分，每個詞彙都是以高維度向量來計算

通常是利用cosine的關係來判斷詞彙的緊密程度

當兩向量的夾角越小，詞彙緊密程度越高

淺談自然語言處理

第一種是「符號表示」，bat在中文裡是蝙蝠，在英文中卻可以有球棒、球拍兩種意思，這就是歧異性。這是要注意的一件重要的事。

一個詞代表的意思特別重要