自然語言處理(台灣大學資訊工程學系 陳信希教授)

1. 自然語言處理的應用和挑戰：
2. 機器翻譯：把A語言翻譯成B語言

挑戰：A、B語言間使用的詞彙和結構有差異

目的：生活上的需求，圖像文字與語音整合翻譯

1. 問答系統：問問題與回答問題

挑戰：分析問題、分析內容、計算支持或反駁資訊的信心度

技術：自然語言處理、資訊檢索、機器學習、知識表示和推理、大規模平行計算

1. 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、病例等不同類型來源的數位資料，進行文件探勘。了解產品資訊銷售額、顧客評價。

用途：尋找旅館、市場分析、選情分析

1. 自然語言處理的典型問題
2. 歧義解析

* 詞彙層次(同一英文單字在不同的詞性上有許多中文翻譯)
* 語法層次(介係詞片語不知道是在修飾哪段話)
* 語意層次

1. 容錯力

* 打錯字
* 語法錯誤
* 未知詞(未被收錄之詞彙，例如：網紅)

1. 強健性

* 領域改變
* 網路語言(縮略詞，例如：不要降=不要這樣；諧音，例如：BJ4=不解釋)
* 表情符號

1. 淺談自然語言處理
2. 概念表示：概念由符號表示，不同語言有不同方式所表示，例如：中文為貓，英文為cat，但符號有歧義性
3. 語言單位、成分或單元：真實世界的概念以符號組合後代表，再組成句子或段落
4. 掌握共通性：分類[自然語言處理]
5. 詞性類別[詞性標記]：名詞、動詞、副詞…
6. 語義類別[語義關係標記]：同義詞(例如：夕陽、斜陽、落日)
7. 句法類別[句法剖析]：名詞片語、介系詞片語…
8. 相依類別[相依剖析]：詞彙與詞彙的分類
9. 言談類別[言談剖析]：時序、因果、轉折、推展
10. 意見類別[意見探勘]：正面與負面的詞彙
11. 情感類別[情感分析]：憤怒、厭惡、恐懼、喜悅、悲傷、驚訝…
12. 立場類別[立場偵測]：贊成與反對某個議題
13. 分佈式表示
14. 意思：同一詞彙在不同語句中可能表達不同的意思，若要了解詞彙的意思，關鍵是伴隨出現的詞彙
15. 分佈式假設：兩個詞的上下文相似，則這兩個詞的詞義也相似
16. 向量計算
17. 分散式表示
18. 向量空間詞彙表示(預測)
19. 低維度稠密向量：語法關係、語意關係
20. 類比運算(\_：\_ = \_：\_，單複數關係、性別關係…)
21. 向量運算(a：a\*=b：b\*)