**人工智慧 – 自然語言查詢**

**系別 : 數位三 姓名 : 林玟廷 學號 : S10755045**

* **機器翻譯:**  
  不同語言間的文本自動翻譯；原始語言和目標語言使用詞彙和結構差異；原始語言文具分析，目標語言文去生成。
* **問題系統挑戰:**  
  分析問題，找出再問甚麼；分析內容，擷取正確答案。
* **自然語言處理的典型問題:**
* 容錯力: 打錯字、語法錯誤、未知詞。
* 強健姓: 領域改變、網路語言(如 : 方言俗語、縮略語、符號合併等)、表情符號。
* **分類 – 掌握共通性:**
* 詞性類別: 普通名詞、地方詞、狀態句賓動詞等。
* 語意類別: 夕陽、斜陽、落日(同義詞詞林)。
* 句法類別: PP、NP、S等。
* 相依類別: nsubj、root、punct、dobj。
* 言談類別: 時序、因果、轉折、推展。
* 意見類別: 正面、負面。
* 情感類別: 憤怒、厭惡、恐懼等。
* 利長類別: 如同性婚姻，贊成、反對。
* **自然語言處理一些任務:**
* 詞素: 詞素分析。
* 詞彙: 中文斷句。
* 命名實體: 命名實體辨識。
* 詞性: 詞性標記。
* 句法類別: 句法剖析。
* 相依類別: 相依剖析。
* 語義類別: 語義關係標記。
* 言談關係: 言談剖析。
* 意見: 意見探勘。
* 情感: 情感分析。
* 立場: 立場偵測。
* **符號計算:**  
  詞彙和類別都是以符後呈現，匹配是基本符號計算，其無法計算詞彙語意關聯程度。
* **分布式表示:**
* 意思: 意思的產生來自使用，要了解詞彙的意思，關鍵是伴隨出現的詞彙。意思相似的詞彙很可能出現在相似的語境。分布式假設: 如果兩個詞的上下文相似，則這兩個詞的詞義式相似。
* 語境: 上下文，共同出現，以向量計算表示。
* **分散式表示:**  
  將詞彙轉換成低微度稠密向量，如: 語法關係、語義關係。
* 類比運算
* 向量運算
* **Word2vec的延伸:**  
  由符號的離散資料呈現，改變成連續資料呈現。
* 語言成分表示: 詞向量、句子向量。
* 成分類別表示: 詞性向量、詞義向量。
* 用戶資料表示: 用戶向量。
* **相似關係計算:**
* 詞彙間相似關係 : 詞彙分群。
* 句子間相似關係 : 句子分群。
* 段落間相似關係 : 段落分群。
* 文件間相似關係 : 文件分群。
* 問句和答句間關係 : 問答系統答案選擇。
* 句子-文件之間關係 : 文件搜尋。