

處理自然語言的簡化法

本院覽號

公告日期

智財權狀態

05A-1081111

2024-01-19

美國臨時案已申請、美國US 11,568,151 B2已獲證、中國已申請、台灣(發明)I 735380已獲證

摘要

一個詞 X 的修飾語，通常是語意上能夠和 X 搭配的詞。一個複雜的句子通常是由簡單句逐步地加上許多語意上適合搭配的修飾語，修飾子句，或者修飾語的修飾語，補語等等。如果我們對每個詞 X 蒐集其修飾語集合FB(X)。就可以利用詞與詞之間的修飾關係，將一個複雜句反推回原來的簡單句。要進行這個計算，我們需要將句子中所有合理的修飾關係利用FB和句子結構推導出來。如此，就會得到這個句子的依存剖析樹。將一個詞 X 的修飾語「併入」X 的動作，我們稱之為「簡化」(reduction)。對一個句子進行簡化，我們從依存剖析樹的端點(leaf node)的詞遞迴地與上面的搭配詞合併，回推至其原來的簡單句。本發明描述一個利用FB產生依存剖析樹的方法，並可同時可進行斷詞以及語言生成。

技術優勢

1. 以往中文斷詞與剖析是分開處理，斷詞的失誤會造成後面剖析系統無可挽回的錯誤。我們的方法讓兩者齊頭並進，同時完成。
2. 提供一個嶄新的專有名詞辨識方法，以概念辭典的方式進行辨識。
3. 可以處理口語及不太合文法的句子。

應用範圍

1. 客服問答、對話系統：幫助理解客戶問句的「意圖」以及回應方式，並進行商品推薦。
2. 機器閱讀：幫助客戶理解網路相關文獻，自動整理醫院病歷至資料庫。
3. 輔助教學系統：扮演電腦家教的角色，協助學生瞭解作業，以及補強學生覺得困難的部分。

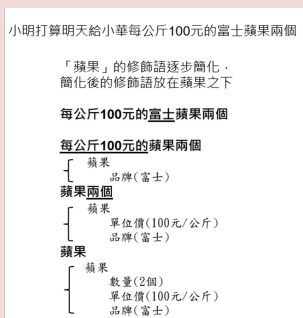


圖1. 蘋果的修飾語簡化的過程

圖1. 蘋果的修飾語簡化的過程

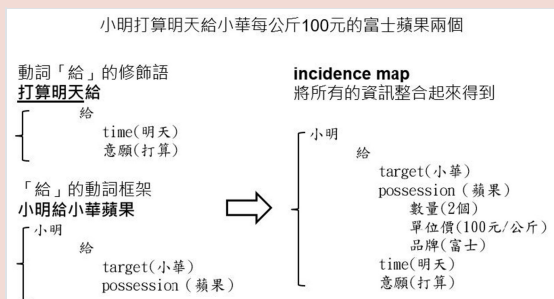


圖2. 整句話的 incidence map

圖2. 整句話的 incidence map

創作人

許聞廉



中央研究院
ACADEMIA SINICA