NLP Literature Review (Loki部分)

Articut/Loki (Wang et al., 2019) 是一個基於生成語言學原理的自然語言處理 (NLP) 和自然語言理解 (NLU) 系統。Loki代表Linguistic Oriented Keyword Interface，它能在保留自然語言結構本質的前提下進行運算。Articut以X-Bar Theory為基礎，執行中文斷詞 (Chinese Word Segment，簡稱CWS) 同時加註詞性標記 (Part of Speech，簡稱POS)，還包括命名實體辨識 (Named Entity Recognition，簡稱NER)，即使用自定義辭典來定義正式詞彙，形成一個完善的NLU系統。

在這個系統中，POS對於斷詞非常重要，它決定了詞性的同時也完成了斷詞。此外，NLU系統需要NER以滿足對正式詞彙（如人名、地名、複合詞等）的需求。語言經過斷詞處理，為計算機提供有意義的單詞，使語言更容易處理。

Loki是根據語言學原理來進行設計，符合現實中語意判讀的產生。也就是在特定**語境**下，用完整或不完整的句子作為**語言表達**，以確認**意圖。**本工具採用了自然語言處理（NLP）領域常見的術語，也將其分為這三個層次：專案名稱（Project Name）、意圖名稱（Intent Name）、語言表達（Utterance）。其中專案名稱正是對應對話中的語境，而Loki的設計確實能做到實現以上概念。

首先，開發者能自行設定專案名稱。根據本研究內容，著重在談論含有中文wh詞「誰」的語境世界。因此設定shei\_sinica (中文wh疑問詞：誰) 作為專案名稱(Project Name)，對應到語意學定義中的語境世界。shei\_sinica專案下的意圖則適用於含有中文wh疑問詞「誰」的語境。接著建立了三種不同「語意」的意圖(Intent)，分別為疑問語意，存在語意和全稱語義，在這些意圖下，所有句子都旨在描述這三種「語意」的語言表達(Utterance)。例如，「無論誰當權」、「誰做好公民誰就是傻瓜」屬於全稱語意意圖之下的不同語言表達形式。

而透過Loki呼叫Articut進行斷詞、詞性標記和命名實體辨識處理，所有句子都轉化為僅包含詞性標記(POS)正規表達式(Regular Expression，簡稱Regex)，並保留其順序作為句法結構，以實現最大的兼容性。在程式端，Loki比對語料結構，並使開發者能夠在相關程式區塊中編寫程式碼，針對不同結構抽取需要的詞彙參數、並自行調整、設計回應。

(緊接建模)~~例如…~~

有了 Loki 意圖模型，符合全稱語義句型的文字，就會被 Loki 分類為 universal\_use意圖。(建模最後一句)