

天主教輔仁大學圖書資訊學系碩士班碩士論文

指導教授：陳舜德 博士

基於自然語言處理技術整合維基百科(Wiki)之
圖書館參考諮詢機器人建置與使用評估

Development and Evaluation of a Library Chatbot
for Reference Service Integrating Natural Language
Processing Based on Wiki

研究生：黃丰嘉 撰

中華民國 110 年 6 月

摘要

自 2016 年起，聊天機器人 (Chatbot) 的風潮崛起且應用遍地開花，舉凡股市理財投資 (Line @微股力 ScanTrader)、旅遊搜尋推薦 (Facebook Messenger @Skyscanner)、披薩訂購 (Google Dialogflow @達美樂 Domino) ... 等皆能看到聊天機器人的身影。聊天機器人的出現，提供使用者客製化、不受時空限制的服務體驗，在大大節省客服人力資源之餘，也提高處理重複性工作的效率。如今，隨著語意分析 (semantic analysis)、機器學習... 等技術的大幅躍進，聊天機器人不再是簡單的「關鍵字觸發」任務，而是從大量語料中透過對字詞、句型、句法結構的分析來理解使用者的詢問目的 (意圖)，也就是所謂的自然語言理解 (NLU, Natural Language Understanding) 技術。

圖書館作為重要的知識典藏機構，應當思考如何運用新興科技提供讀者一個更加行動化、個人化與智慧化的參考諮詢介面，從而發展出一套便捷、友善、效能的服務模式，是個值得期待的嘗試。因此，本研究以圖書館為應用場景、LINE 為圖書館線上參考諮詢的系統介面，實際建置一個解答館務問題和提供學科知識服務的參考諮詢機器人—FJCU Lib DialogSys。有別於「規則式」的制定問句的問法，本研究的貢獻在於讓讀者與系統之間的對話得以彈性，經由下列三種方式達成：其一，利用「維基百科 (Wikipedia)」更新即時、詞彙描述淺顯多樣以及內容豐富且持續成長... 等特性，協助問句進行查詢擴展 (Query Expansion)、作為斷詞 (Word Segmentation) 程序的擴增語彙集，以及用以回應讀者學科知識諮詢。其二，在問句的處理上，運用「自然語言處理 (NLP, Natural Language Processing)」技術來幫助判斷讀者的詢問意圖，進而獲得切合其資訊需求的答覆。其三，在界定詢問意圖的顆粒度方面，本研究使用基於密度的 DBSCAN 分群演算法藉由參數的調整將具有類似問法的訓練語料歸為同一個意圖，並儲存於 MongoDB 內。而後，若有增加其他意圖的必要性時，也能輕易地擴充訓練語料，有助於後續的維護與利用。

關鍵字：意圖式聊天機器人、自然語言處理、Wikipedia、查詢擴展、特徵擷取、DBSCAN

Abstract

Since 2016, the trend of Chatbot has risen and applications have blossomed everywhere, such as stock market financial investment (Line @微股力 ScanTrader), travel search recommendations (Facebook Messenger @Skyscanner), pizza ordering (Google Dialogflow @達美樂 Domino), etc. The emergence of Chatbot provides users with a customized service experience that is not limited by time and space, which greatly saves customer service human resources and improves the efficiency of handling repetitive tasks. Nowadays, with the rapid advancement of semantic analysis, machine learning, and other technologies, chatbots are no longer "keyword trigger" tasks, but through matching words, sentence patterns, and syntax from a large amount of corpus. The analysis of the structure to understand the user's inquiry purpose (intention), which is called natural language understanding (NLU) technology.

As an important knowledge collection institution, libraries should think about how to use emerging technologies to provide readers with a more mobile, personal, and intelligent reference interface, to develop a set of convenient, friendly, and efficient service models, which is worth looking forward to trying. Therefore, this research uses the library as the application scenario and LINE as the library's online reference consultation system interface and builds a reference consultation robot — FJCU Lib DialogSys, which answers library FAQs questions and provides subject knowledge services. Different from using the "Rule-based" method to set keywords to trigger reply, the contribution of this research is to make the dialogue between the reader and the system flexible, which can be achieved through the following three ways: First, use "Wikipedia" because of its quickly updating, diverse vocabulary description, rich in content and continuous growth, etc., which not only assist in the query expansion of the FAQ questions, as a extend word set for the Word Segmentation preprocessing, but also to respond to readers' subject knowledge inquiries. Second, in the processing of FAQ questions, "NLP (Natural Language Processing)" technology is used to help judge the reader's intent to obtain a response that meets their information needs. Third, in terms of defining the granularity of query intent, this study uses the density-based DBSCAN clustering algorithm to group training corpora with similar question structures into the

same intent through parameter adjustment and store them in MongoDB. Then, if it is necessary to add other intentions, the training corpus can also be easily expanded, which is helpful for subsequent maintenance and utilization.

Keyword: Intent-based Chatbot, Natural Language Processing, Wikipedia, Query Expansion, Feature Extraction, DBSCAN

