國 立 清 華 大 學

碩 士 論 文

使用基於生成式 AI 之智慧工程文件摘要系統以達成流程優化之成果

Using an intelligent engineering document summarization system based on generative AI to achieve process optimization

系別：工業工程與工程管理學系 組別：工業管理組

學號: 111034602

研 究 生： 林憶姍 (Yi-Shan Lin)

指導教授：張瑞芬博士 (Prof. Amy J.C. Trappey)

中 華 民 國 一一三 年 七 月

摘要

工程邀標書(Request for Quotation, RFQ)為各式工程進行招標時，發布給具製造能力之潛在投標廠商的文件，內容包含了項目要求、技術規範、合約條款、預算……等重要資訊，大型變壓器製造等高度客製化工程產業，經常使用 RFQ 文件作為進行詢報價的工具。RFQ 蘊含之資訊內容繁瑣，有許多不可疏漏之關鍵資訊，且因高度客製化產業之顧客規範的需求複雜度高，需耗費工程技術專業人員許多的心力及時間閱讀並截取重點，以進行成本估算而後產出最佳報價。但招標通常有時間的限制，要在短時間內完成整體報價流程極具難度，且產品設計之報價多是透過專業人員的經驗法則而得，每位人員判斷基準皆有所不同，會產生報價結果不一致的狀況。因此，本研究期望能透過 RFQ 文件的自動摘要系統，讓報價人員可以更快速的將變壓器關鍵規格表填答完畢，並提出較一致且合理的報價，藉此提高企業的資源使用率及資訊準確性。本研究利用變壓器規格的關鍵字，提取出 RFQ 全文中含有關鍵字的重要句子，做為 RFQ 訓練集之摘要，並在經由資料集的驗證後，獲得約 2,252 筆前案 RFQ 文件集與規格原文、摘要對照之資料組合作為模型的完整訓練集(Training datasets)，並對 Transformers 架構的 PEGASUS 自然語言處理模型進行微調(Fine-tune)，提升模型產出結果和語法的精準度。接下來，是利用透過規格表填答率與 ROUGE 摘要品質評估方法來進行文本內容及摘要的比對和評估，並經由工程領域專家的驗證及電力變壓器規格表的完整性，來確認摘要結果的正確性和有效性，以組成一產生工程文件專用之自動化摘要生成流程。最後，本研究使用新 RFQ 進行驗證整體流程的驗證，期望可增加成本評估與報價的流暢度。

關鍵字: 關鍵字提取、自然語言處理、自動化摘要

Abstract

The Request for Quotation (RFQ) is a document issued to potential manufacturing bidders when various engineering projects are open for bidding. It encompasses crucial information such as project requirements, technical specifications, contract terms, budget, and other vital details. In industries with highly customized projects, like the manufacturing of large transformers, RFQs are commonly used as tools to solicit price quotations. The information within an RFQ is intricate, containing numerous critical details that must not be overlooked. Due to the complexity of customer specifications in highly customized industries, such as the production of large transformers, reading and extracting key information from RFQs demand significant effort and time from engineering professionals. The bidding process often has tight deadlines, making it challenging to complete the entire quotation process within a short timeframe. Moreover, the pricing of product designs typically relies on the experiential judgment of professionals, leading to inconsistencies in quotation results due to differing individual assessment criteria. Therefore, this research aims to streamline the process by developing an automatic summarization system for RFQ documents. This system enables quotation professionals to quickly fill in key specifications for transformers, resulting in more consistent and reasonable quotations. The study involves extracting important sentences containing keywords related to transformer specifications from the entire RFQ text to create a summary, using these summaries as a training dataset. After validation, approximately 2,252 records of previous RFQ documents, along with the original specifications and corresponding summaries, are collected for the complete training set. The PEGASUS natural language processing model based on the Transformers architecture is fine-tuned to improve the precision of the model's output and syntax. The research then assesses the quality of the generated summaries and the

content of the text using the specification completion rate and the ROUGE summary quality evaluation method. Verification by domain experts in the engineering field and confirmation of the completeness of power transformer specifications ensure the correctness and effectiveness of the summary results, forming an automated summarization generation process specifically for engineering documents. Finally, the research validates the entire process using a new RFQ to enhance the fluency of cost estimation and quoting.

Keywords: Keyword Extraction, Natural Language Processing (NLP), Automated Summarization