

中原大學  
工業與系統工程學系  
碩士學位論文

基於 TRIZ 標記專利優化資料庫之研究

Labeling Patent Database Using TRIZ for Optimization

指導教授：劉天倫

研究生：陳崇恩

中華民國 114 年 7 月

## 摘要

本研究旨在結合 TRIZ 理論與自然語言處理技術，在不對人工智慧模型進行微調或重新訓練的狀況下，優化專利資料庫的檢索效能，提升產品開發過程中的創新能力。研究以 TRIZ 中 D. Mann 所提出的 48 種工程參數為基礎，運用預訓練的 BERT 模型對專利文本進行語意分析與特徵標記，建立專利文獻與工程參數間的對應關係。接著，整合由 Facebook 提出的 Retrieval-Augmented Generation (RAG) 架構，將使用者輸入的工程問題轉換為語意查詢，從資料庫中快速檢索出與問題高度相關的專利文本，提供創新解決方案的參考依據。本研究不僅提升專利檢索的準確性與效率，降低傳統查詢所需的時間與成本，也不需要對任何人工智慧模型進行微調，大幅降低訓練人工智慧模型的成本，並促進智慧型創新輔助系統的發展。研究成果顯示，本系統能有效協助使用者解決工程問題，提升專利資料庫在研發流程中的實用性，實現以知識驅動創新的目標。

**關鍵字：**自然語言處理、BERT、RAG、專利檢索、創新輔助工具

中 原 大 學

## Abstract

This study aims to integrate TRIZ theory with natural language processing (NLP) techniques to enhance the retrieval efficiency of patent databases and improve innovation in product development, while avoiding any fine-tuning or retraining of artificial intelligence (AI) models. Based on the 48 engineering parameters proposed by TRIZ theory, the study uses a pre-trained BERT model to perform semantic analysis and feature annotation on patent texts, thereby establishing a mapping between patents and engineering parameters. Furthermore, the study integrates with Facebook's Retrieval-Augmented Generation (RAG) framework to transform user-input engineering problems into semantic queries, enabling rapid retrieval of highly relevant patent documents as references for innovative solutions. This approach not only improves the accuracy and efficiency of patent searches and reduces the time and cost associated with traditional query methods but also reduces the need for fine-tuning AI models while lowering the cost of AI model training. The results demonstrate that the proposed system effectively assists users in solving engineering problems and enhances the practical value of patent databases in R&D processes, ultimately practicing knowledge-driven innovation.

**Keywords :** NLP, TRIZ, BERT, RAG, Patent Retrieval, TRIZ-based Design Tools

## 致 謝

首先，向我的指導教授劉天倫老師致上最深的感謝。在整個研究過程中，劉老師不僅在研究方向的規劃、研究方法的選擇、以及論文內容的撰寫上提供了許多啟發性的建議，更以其豐富的學識與嚴謹的學術態度，引領我不斷思考與精進，使本研究得以更加完善。此外，感謝老師提供我參與國際發明學會研討會（ICKII）的機會，讓我可以與來自各地的優秀研究者進行學術交流，拓展我的視野，也豐富了我的人生經歷。老師也在本論文的完成過程中持續提供寶貴意見與修正方向，對我而言受益匪淺。

在研究所這段期間，我也要感謝研究所同學明政、凌翔、江謙、廷軒、柏群、乙萱、昌錫、子奕、珮渝與一輝的協助。感謝他們在資料蒐集、助教工作及論文瓶頸時，給予我幫助，並在文獻整理、論文撰寫方式上提供許多實用建議。特別感謝珮渝同學在論文題目的發想與方向上給予我啟發與靈感，也感謝一輝同學建置了龐大的專利資料庫，讓我在整理與收集原始資料的過程中節省了大量時間與精力。

最後，感謝我的家人與朋友，尤其是我的父母，感謝他們在我研究期間無私地給予我生活上的照顧與精神上的支持。在我遇到困難與挫折時，他們總是默默陪伴與鼓勵，成為我堅持下去的動力，讓我能夠持續投入研究，並順利完成本論文。

陳崇恩 謹誌於

中原大學 工業與系統工程學系

中華民國一一四年月

# 目 錄

摘 要 .....	I
Abstract .....	II
致 謝 .....	III
目 錄 .....	IV
圖 目 錄 .....	VI
表 目 錄 .....	VII
第一章 緒 論 .....	1
第一節 研究動機 .....	1
第二節 研究目的 .....	3
第三節 研究流程 .....	4
第二章 文獻探討 .....	5
第一節 創新發明問題解決理論 .....	5
第二節 48 種工程參數 .....	6
2.2.1 48 種工程參數 .....	7
第三節 自然語言處理 .....	10
2.3.1 簡介 .....	10
2.3.2 TF-IDF .....	10
2.3.3 SBERT .....	10
第四節 多元分類器 .....	12
第五節 RAG .....	13
2.5.1 檢索器 .....	14
2.5.2 向量資料庫 .....	14
2.5.3 LangChain 工具 .....	15
第六節 專利探勘 .....	16
第三章 研究方法 .....	17
第一節 專利文本處理 .....	19
第二節 48 種工程參數字詞庫建立 .....	21
第三節 標籤專利文本流程 .....	21
3.3.1 各分類器分數計算方式 .....	22
3.3.2 工程參數分數計算流程 .....	22
3.3.3 專利文本標記方法 .....	23

第四節 應用檢索器 .....	24
3.4.1 向量資料庫索引化 .....	24
3.4.2 檢索向量資料庫 .....	25
第五節 分析結果流程 .....	26
第四章 研究結果 .....	27
第一節 專利文本標籤的結果 .....	27
第二節 檢索器結果分析 .....	32
4.2.1 檢索效率驗證 .....	32
4.2.2 檢索結果比較 .....	34
4.2.3 檢索結果文本分析 .....	36
第五章 結論與建議 .....	39
第一節 結論 .....	39
第二節 研究限制 .....	39
5.2.1 關鍵字詞庫的建立 .....	39
5.2.2 分類器語文本權重 .....	40
5.2.3 檢索器設計 .....	40
5.2.4 標籤與檢索結果驗證 .....	40
第三節 未來研究與建議 .....	40
參考文獻 .....	42
附錄二 48 工程參數關鍵字詞表 .....	50
工程參數代號 .....	50
關鍵字詞 .....	50

## 圖目錄

圖 1 - 1、研究流程圖 .....	錯誤! 尚未定義書籤。
圖 2 - 1、TRIZ 解題流程 .....	5
圖 2 - 2、siamese 架構圖(Reimers et al. 2019) .....	11
圖 2 - 3、SBERT 推論架構圖(Reimers et al. 2019) .....	錯誤! 尚未定義書籤。1
圖 2 - 4、Polikar 提出的其中一種決策架構 (Polikar, 2006) .....	錯誤! 尚未定義書籤。2
圖 2 - 5、RAG 架構流程(Gao et al. 2024) .....	錯誤! 尚未定義書籤。3
圖 2 - 6、RAG 詳細架構 (Lewis et al. 2020) .....	錯誤! 尚未定義書籤。3
圖 2 - 7、USPTO 專利文本資訊 .....	錯誤! 尚未定義書籤。6
圖 3 - 1、研究方法流程圖 .....	錯誤! 尚未定義書籤。8
圖 3 - 2、提取欄位後的資料 .....	錯誤! 尚未定義書籤。9
圖 3 - 3、原始文本內容 .....	錯誤! 尚未定義書籤。9
圖 3 - 4、進行處理後的文本內容 .....	錯誤! 尚未定義書籤。0
圖 3 - 5、處理後資料概觀 .....	錯誤! 尚未定義書籤。0
圖 3 - 6、48 種工程參數關鍵字詞庫(部分) .....	錯誤! 尚未定義書籤。1
圖 3 - 7、專利文本工程參數分數表 .....	錯誤! 尚未定義書籤。
圖 3 - 8、索引化步驟 (Chase) .....	錯誤! 尚未定義書籤。4
圖 3 - 9、檢索器虛擬碼說明 .....	錯誤! 尚未定義書籤。
圖 3 - 10、檢索器虛擬碼說明 .....	錯誤! 尚未定義書籤。5
圖 3 - 11、檢索器虛擬碼說明 .....	錯誤! 尚未定義書籤。6
圖 4 - 1 查詢語句字詞匹配數 .....	錯誤! 尚未定義書籤。8

中原大學

## 表 目 錄

表 2 - 1、48 項工程參數.....	6
表 3 - 1、查詢語句表.....	錯誤! 尚未定義書籤。6
表 4 - 1、五篇專利文本編號與標題.....	錯誤! 尚未定義書籤。7
表 4 - 2、專利編號 14539847 分析表.....	錯誤! 尚未定義書籤。7
表 4 - 3、專利編號 12385303 分析表.....	錯誤! 尚未定義書籤。8
表 4 - 4、專利編號 13794125 分析表.....	349
表 4 - 5、專利編號 12397463 分析表.....	錯誤! 尚未定義書籤。
表 4 - 6、專利編號 12393670 分析表.....	錯誤! 尚未定義書籤。
表 4 - 7、檢索花費時間比較表.....	錯誤! 尚未定義書籤。3
表 4 - 8、檢索文本最高相關專利文本比較.....	錯誤! 尚未定義書籤。4
表 4 - 9、檢索工程參數結果表.....	錯誤! 尚未定義書籤。6



中 原 大 學



# 第一章 緒論

近年來科技日新月異，人工智慧所帶來的便利性讓產品創新的速度不斷加快。稍有停滯，便可能在市場中落於人後。因此，如何打造兼具創新性與市場需求的新產品，成為企業保持競爭力的關鍵，而持續創新更是一項極具挑戰性的課題。

現今的產品設計日益複雜，顧客期望也不斷提升，使得新產品開發面臨重重瓶頸。突破現狀、維持創意與差異化，成為企業不可避免的課題。在開發新產品過程中，專利檢索是一項常見的方式，目的在於激發創新靈感。根據經濟部統計處統計，2023 年的專利申請量高達 7.2 萬件，顯示專利資訊的龐大與重要性。

雖然現今已有 ChatGPT、Copilot、Gemini 等大語言模型(LLM)可協助檢索專利資料，並提供建議，但這些通用模型多半缺乏專利領域的專業知識，加上無法有效處理大量資料，導致回覆品質不一，難以為開發者帶來實質幫助。

為了解決此問題，雖然可以透過微調(fine-tuning)或訓練專用語言模型來提升準確性與效能，但此過程不僅需要高階硬體設備與大量時間準備資料與訓練，很多研究也表示，訓練專屬模型或進行微調可能會導致知識退化、幻覺風險、數據偏差等結果。加上語言模型推陳出新的速度也愈快速，開發專屬模型在時間與經濟效益上顯得不理想。

因此，如何在不重新訓練語言模型的前提下，有效改善現有模型的應用方式，能在檢索巨量專利文本下，協助產品開發者尋找創新解方的一大課題，需要深入探索與解決。

## 第一節 研究動機

在新產品開發的過程中，常會面臨各種技術矛盾。例如，為提升手機的觀看體驗，螢幕尺寸必須增大，但過大的螢幕又會降低裝置的攜帶便利性。這類問題可透過 TRIZ 工具中的矛盾矩陣，找出對應的發明原則並提出解決方案，摺疊手機即為其中的一種解決方案。類似的技術矛盾在不同產業領域中屢見不鮮，雖然可能已有解方，但往往散佈於不同專利中。

目前 TRIZ 的應用仍高度依賴專家的經驗與判斷。專家在閱讀專利文本時，容易因主觀看法差異而忽略潛在的創新線索，加上分析專利內容所耗費的時間也相當可觀。因此，若能結合自然語言處理（NLP）技術與有效的檢索方法，將有助於降低人工分析成本，並提供更客觀且廣泛的資訊參考。

儘管現有的大語言模型具備強大的語言理解能力，但在專利分析領域仍存在幾項挑戰。一方面，多數模型缺乏專利資料庫的來源，即使具備檢索功能，面對龐大的資料量仍易導致效能低落。另一方面，現有模型對 TRIZ 知識掌握有限，導致回答品質參差不齊。雖然可透過微調或訓練專屬模型來改善，但研究顯示（如 Gekhman et al., 2024; Ghosh et al., 2024），微調過程可能引發模型幻覺、知識退化與風格化學習等問題。此外，Bender et al. (2021) 與 Malikeh Ehghaghi (2024) 指出大型語言模型訓練成本高昂且容易強化偏見或學習表面形式。

相較之下，Lewis et al. (2020) 提出的 RAG (Retrieval-Augmented Generation) 架構，透過外部檢索的機制與生成模型，有效補足語言模型的知識盲點，並降低幻覺與偏見風險。因此，若能結合 RAG 架構進行專利文本的檢索與分析，加上基於 TRIZ 對專利文本進行標記，可以幫助產品開發者於專利檢所與產品上的創新。

## 第二節 研究目的

上述僅為概念性架構，實際上，如何標註專利文本並整合至 RAG 檢索架構，以建立智慧化的推薦流程，仍是極大的挑戰。根據 Polikar (2006) 的研究，採用多元分類器並整合結果與權重，有助於在資料過多或不足的情境下提升分類準確性，並可進一步支援增量式學習。蔡元豪 (2012) 提出透過建立關鍵字詞庫並結合 TF-IDF 權重進行文本分類，有助於提升語意識別精度。而 Tseng et al. (2014) 則發現，TRIZ 中的工程參數比發明原則更適合作為分類基礎，有助於更準確地反映技術特性。此外，Reimers et al. (2019) 所提出的 SBERT 架構，使得 BERT 模型能夠快速產生語句嵌入，進而支援語意相似度計算與語句層級的檢索任務。

綜合上述研究成果，本研究可採取以下步驟以構建專利推薦系統，首先建立以 TRIZ 工程參數為基礎的關鍵詞庫。其次，將專利文本切分為語意完整的段落。接著，採用多種分類器進行段落標註，並以權重進行整合。最終，導入 RAG 檢索架構以執行專利資料庫檢索，從而提供高相關性的專利文本推薦，達成檢索效率的優化，並作為創新解方的輔助工具。

### 第三節 研究流程

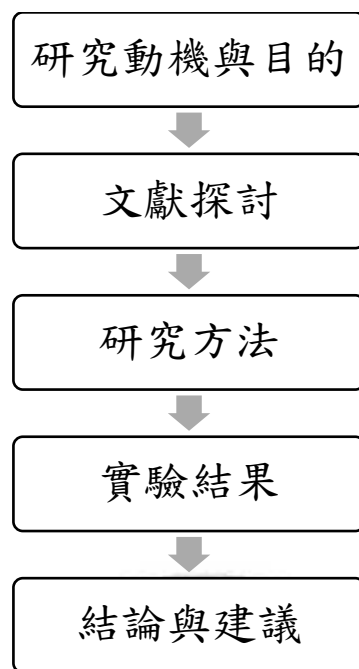


圖 1-1、研究流程圖

本研究的論文架構與流程如上圖 1-1 所示，為了達到專利文本的標註方式以及整合 RAG 架構，此研究分成五個部分：

1. 第一部分：本段旨在探討 TRIZ 理論與大型語言模型的發展現況，分析其優勢與局限性，並說明本研究試圖解決的問題。
2. 第二部分：回顧並彙整相關領域之研究，包括 TRIZ 理論與工程參數、自然語言處理技術(NLP)、語句表示模型(SBERT)、多元分類器架構、以及 RAG 檢索架構，作為本研究方法的基礎。
3. 第三部分：將詳述整體資料流程與輔助創新架的架構，包含專利文本資料來源、資料前處理流程、工程參數關鍵字詞庫建立、文本標註方法、多分類器加權架構、RAG 架構導入方式。
4. 第四部份：本部分針對系統所產出的推薦結果進行評估，透過量化的指標與分析，探討其表現。
- 第五部分：總結本研究成果與研究限制，並提出後續可能延伸之研究方向與建議，以進一步完善本智慧化輔助工具之實用性。

## 第二章 文獻探討

### 第一節 創新發明問題解決理論

TRIZ 是俄文首字母縮寫，英譯為 Theory of Inventive Problem Solving，是由已故的科學家 Genrich Altshuller 及其團隊，在分析大約 40,000 項技術專利後，於 1946 年所提出的解決發明問題理論。Altshuller 有系統地歸納問題並提供分析工具與方法，對產品上的創新設計與發想有很大的幫助。在簡毓汝 (2004) 中證明其有效性與可用性。

TRIZ 理論有五個基本概念，「功能」、「資源」、「矛盾」、「理想」、「空間-時間-介面」。基於上述這些概念，延伸出：問題階層分析 (Problem class analysis)、功能分析 (Functional Analysis)、矛盾矩陣表 (Contradiction Matrix)、四十項創新法則 (40 Inventive Principles)、演化趨勢 (Trends)、物質-場理論 (Su-Field)、最終理想化結果 (Ideal Final Result, IFR)、科學效應 (Effects)、TRIZ 演算法則 (Algorithm for Inventive-Problem Solving)、因果鏈分析 (Cause-Effect Chain Analysis, CECA) 等背後人所建立的工具，可以說 TRIZ 從提出以來，發展得相當成熟，也被應用於工程、管理、教育、醫療、財務、客戶服務、品質管理、未來技術以及商業機會辨識等 (蔡明政，2021)。

TRIZ 提供一套從問題定義到問題解決的系統性流程，如圖 2-1。先將特定問題轉化成 TRIZ 一般問題，再產生一般解，最後再找出特定解。

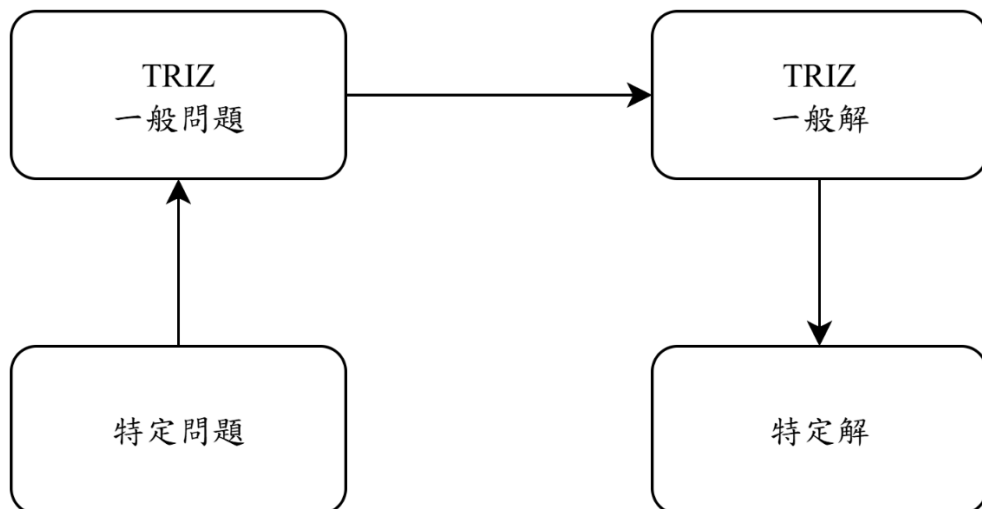


圖 2-1、TRIZ 解題流程

## 第二節 48 種工程參數

原本 Altshuller 提出了 39 種工程參數，所列出的工程參數是在開發產品中遇到的技術或物理矛盾有關。但列出的這些工程參數，仍無法充分解決某些問題，因此於 2003 年時，D. Mann 在 Matrix 2003 中，使用更精確地描述並將工程參數擴展至 48 種，如表 2-1 所示。

表 2-1、48 項工程參數

1. Weight of moving object	2. Weight of stationary object
3. Length or angle of moving object	4. Length or angle of stationary object
5. Area of moving object	6. Area of stationary object
7. Volume of moving object	8. Volume of stationary object
9. Shape	10. Amount of substance
11. Amount of information	12. Duration of action of moving object
13. Duration of action of stationary object	14. Speed
15. Force or torque	16. Energy used by moving object
17. Energy used by stationary object	18. Power
19. Stress or pressure	20. Strength
21. Stability	22. Temperature
23. Illumination intensity	24. Function efficiency
25. Loss of substance	26. Loss of time
27. Loss of energy	28. Loss of information
29. Noise	30. Harmful emissions
31. Other harmful effects generated by system	32. Adaptability or versatility
33. Compatibility or connectability	34. Ease of operation
35. Reliability or robustness	36. Repairability
37. Security	38. Safety or vulnerability

39. Aesthetics or appearance	40. Other Harmful Effects Acting On the System
41. Manufacturability	42. Manufacture precision or consistency
43. Automation	44. Productivity
45. System complexity	46. Control complexity
47. Ability to detect or measure	48. Measurement precision

### 2.2.1 48 種工程參數

以下為謝凌翔 (2021) 根據 D. Mann (2003) 提出的 48 種工程參數的定義說明與解釋，英文版解釋於附錄一中：

1. 移動物體重量(Weight of moving object)：移動物體施加的品質或重力作用於它的支撐或懸吊。
2. 靜止物體重量(Weight of stationary object)：靜止物體的品質或重力作用於它的支撐或懸吊。
3. 移動物體長度/角度(Length/Angle of moving object)：移動物體的任意一維尺寸，長度、寬度、高度或角度。
4. 靜止物體長度/角度(Length/Angle of stationary object)：靜止物體的任意一維尺寸，長度、寬度、高度或角度。
5. 移動物體面積(Area of moving object)：與移動物體的表面或表面積的內部或外部相關的任意二維尺寸。
6. 靜止物體面積(Area of stationary object)：與靜止物體的表面或表面積的內部或外部相關的任意二維尺寸。
7. 移動物體體積(Volume of moving object)：與移動物體或周圍空間所包含到的三維尺寸。
8. 靜止物體體積(Volume of stationary object)：與靜止物體或周圍空間所包含到的三維尺寸。

9. 形狀(Shape)：系統或物體的外觀或輪廓。
10. 物質的數量(Amount of substance)：物體的元件數量或物質總量。
11. 訊息的數量(Amount of information)：物體或系統內的資料或訊息數量。
12. 移動物體作用時間(Duration of action of moving object)：移動物體可執行動作時，所需用時或失效之前的服務壽命。
13. 靜止物體作用時間(Duration of action of stationary object)：靜止物體可執行動作時所需用時或失效之前的服務壽命。
14. 速度(Speed)：物體的速度或任何類型的過程或動作的速率，包含相對或絕對速度、線性或旋轉速率。
15. 力(Force/Torque)：指在改變物體狀態或形態時的任何互動或影響，亦適用扭矩，包括靜態力與動態力。
16. 移動物體消耗能量(Energy used by moving object)：移動物體作用期間所需的能量消耗。
17. 靜止物體消耗能量(Energy used by stationary object)：不動物體在作用期間所需的能量消耗。
18. 功率(Power)：能量使用的速率，指功與時間的比率。
19. 張力/壓力(Stress/Pressure)：單位面積所受的力，指力對物體的影響。
20. 強度(Strength)：物體本身抵抗破壞或耗損的能力。
21. 結構穩定性(Stability)：系統或物體面對內外部因素引響時，其內部結構抵抗互動所產生的改變能力。
22. 溫度(Temperature)：系統或物體的溫度狀態，其中包括各項熱力參數。
23. 光度(Illumination intensity)：單位面積的光通量以及各項亮度特性。
24. 運行效率(Function efficiency)：物體或系統達到主要功能或相關功能的完全執行能力。
25. 物質浪費(Loss of substance)：對系統動作並無貢獻的物質消耗。
26. 時間浪費(Loss of time)：系統在完成指定動作時所額外增加的時間耗損。
27. 能源浪費(Loss of energy)：系統在執行操作時消耗的能量，且此能量消耗對系統並無



貢獻。

28. 資訊遺漏(Loss of information)：資料或系統輸入項的缺失或丟失。
29. 噪音(Noise)：噪音或相關數據有關的事物，例如頻率、音調和分貝等。
30. 有害的散發(Harmful emissions)：系統或物體額外產生任何形式的污染物，並向周遭擴散。
31. 有害的副作用(Other harmful effects generated by system)：系統或物體內部造成的有害作用，主要補充於參數 24 至 30 之間沒有定義的有害作用。
32. 適應性(Adaptability/Versatility)：指一系統或物體套用到不同系統或環境時的適應性。
33. 相容性/連通性(Compatibility/Connectability)：該一系統在與其他系統連結的程度，是否會產生矛盾或衝突。
34. 易用性(Ease of operation)：系統於操作或使用上的容易程度。
35. 可靠性(Reliability/Robustness)：物體或系統能夠正常執行功能的能力，及不易故障。
36. 可修復性(Repairability)：物體或系統在故障後，可容易進行維修恢復之功能。
37. 安全性(Security)：物體或系統保護自身安全不受外在因素侵害的能力，如入侵、盜用等不利因素。
38. 不易受傷性(Safety/Vulnerability)：指一物體或系統保護因外部因素所造成耗損的能力；一個物體或系統保護自己或其他用戶在操作時不受到傷害的能力。
39. 美觀(Aesthetics/Appearance)：物體或系統在視覺上的體驗優劣。
40. 外來的有害因素(Other harmful effects acting on the system)：及系統因為外部因素造成內部元件品質損害或系統本身效率降低的因素多寡。
41. 製造能力(Manufacturability)：物體或系統在製造過程中的難易度。
42. 製造精度/連貫性(Manufacture precision/Consistency)：製造完成的物品與設計規格一致的程度，精度愈高誤差愈低。
43. 自動化水準(Automation)：系統或物體在無人操作的情況下完成任務的能力。
44. 生產力/生產率(Productivity)：指在單位時間裡系統完成操作或執行功能的次數，及每單位時間完成多少次(個)。

45. 系統複雜性(System complexity): 指形成物體或系統元件的數量和相異性。
46. 控制複雜性(Control complexity): 指用於量測或監控系統之元件的數量和相異性。
47. 測量能力(Ability to detect/Measure): 指系統在進行測量工作的複雜、耗時; 測量困難與精度高。
48. 量測精度(Measurement precision): 指一物體或者系統的實際值與量測值之間的誤差。

## 第三節 自然語言處理

### 2.3.1 簡介

自然語言處理，英譯為 Natural Language Processing (NLP)，是一種讓電腦能夠「理解」、「生產」甚至「回答」人類語言的技術，透過數學模型與演算法理解人類的語言，在深度學習的幫助下，使得自然語言有了突破性的成果，於 2017 年，Vaswani. et al. (2017) 提出了 Transformers 模型，再隔年由 Google Devlin et al. (2018) 以 Transformers 架構發表了 BERT 模型，其雙向訓練來建構語義的概念，對 NLP 產生革命性的突破。

### 2.3.2 TF-IDF

TF-IDF (Term Frequency - Inverse Document Frequency) 主要依據詞彙的出現頻率與反轉文件頻率進行詞向量來理解文本內容 (Sparck Jones, 1972)。蔡明政 (2021) 使用此方法從文本中提取特徵，因此 TF-IDF 具備解釋性與計算效率快的特性，但在處理語意時表現有限，如無法判斷「car」與「vehicle」間的語意相近性。

Salton et al., (1975) 進一步優化 TF-IDF 結合了餘弦相似度的計算，用於資料檢索系統中，現今仍廣泛用於搜尋引擎、文件比對與語意分析任務中。

### 2.3.3 SBERT

SBERT 又可以稱為 Sentence-Transformers，由 Reimers et al. (2019) 所提出。改良了 BERT 模型，使用了 siamese 與 triplet 架構對 BERT 編碼後的語句向量進行微調，使其可直接用於計算句子間的語意相似度。將語句嵌入轉換為固定維度的語意向量後，可進

行高效的向量比對與語意檢索，相較於原始 BERT 模型必須逐一比對語句的方式，大幅降低了推論時間與計算資源。圖 2-2 為 siamese 網路架構圖，圖 2-3 為 SBERT 推論架構圖。

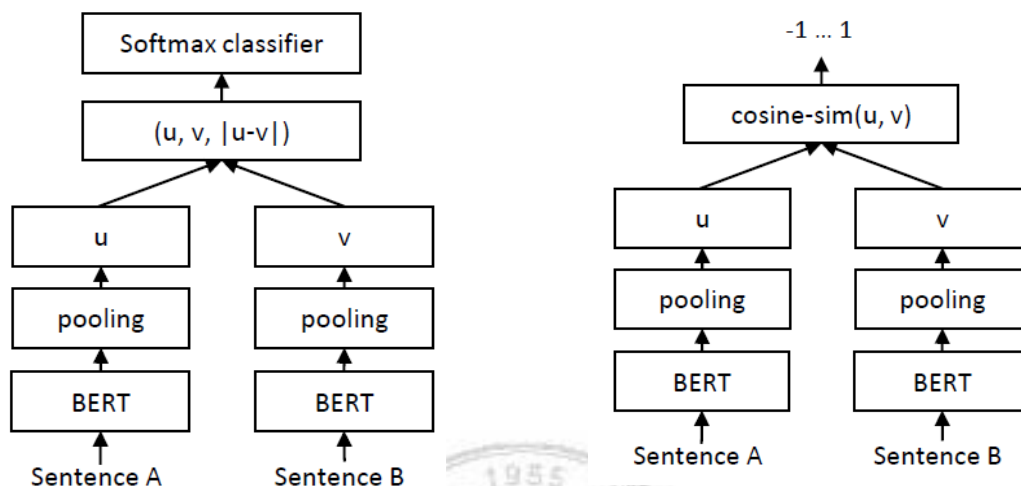


圖 2-2、siamese 架構圖(Reimers et al. 2019)

圖 2-3、SBERT 推論架構圖(Reimers et al. 2019)

在 SBERT 的模型中，設計了分類、回歸以及三元組(triplet)三種目標函數。以下是三種目標函數及其數學公式進行說明：

1. 分類目標函數：將兩句嵌入向量  $u$  與  $v$ ，以及其差的絕對值  $|u - v|$  進行串接並與可訓練的權重  $W_t \in \mathbb{R}^{3n \times k}$  相乘，如公式(1)：

$$o = \text{softmax}(W_t(u, v, |u - v|)) \quad \text{公式(1)}$$

其中  $n$  表示句子嵌入的維度， $k$  為標籤數量，優化了損失函數 (cross-entropy loss)。

2. 回歸目標函數：計算兩個句子嵌入向量  $u$  與  $v$  間的餘弦相似度，並以均方差 (mean-squared-error loss, MSE) 作為目標函數。該方法通常用於語意相似度分數預測，架構如圖 2-3 所示。
3. triplet 目標函數：給定一組錨點句子  $a$ 、正樣本  $p$  及負樣本  $n$ ，透過三元組損失函數(triplet loss) 使  $a$  與  $p$  之間的距離小於  $a$  與  $n$  的距離，並最小化以下的損失函數，如公式(2)：

$$\max (\|S_a - S_p\| - \|S_a - S_n\| + \epsilon, 0) \quad \text{公式(2)}$$

其中  $S_x$  分別為  $a/n/p$  句子的嵌入向量， $\|\cdot\|$  為距離度量， $\epsilon$  為邊界值，確保  $S_p$  比  $S_n$  更接近  $S_a$  至少  $\epsilon$  以上。如此，就可以微調模型並產生具有語意的句子嵌入向量(embedding vector)並進行餘弦相似度計算，有效的提升語意表示及判別性。

### 2.3.4 餘弦相似度

餘弦相似度是自然語言處理中，常使用來衡量向量的方法，會被廣泛運用的原因是因為不受向量長度影響，適合處理不同長度的文本，以表示彼此之間的語意相關性。其公式如下，其中  $A$ 、 $B$  代表不同的文本：

$$\text{sim}(A, B) = \cos \theta = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad \text{公式(3)}$$

## 第四節 多元分類器

Polikar (2006) 系統性地探討了在資料處理、分類與學習策略中，如何透過多個分類器的結合來提升系統整體效能。他指出相較於單一分類器，多個分類器組合能有效降低泛化誤差，處理特徵資料分佈複雜或不均的問題，並提出多種衡量指標與組合策略，有效促進多模型在多維資料中的應用效益。圖 2-4 為提出架構的其中一種。

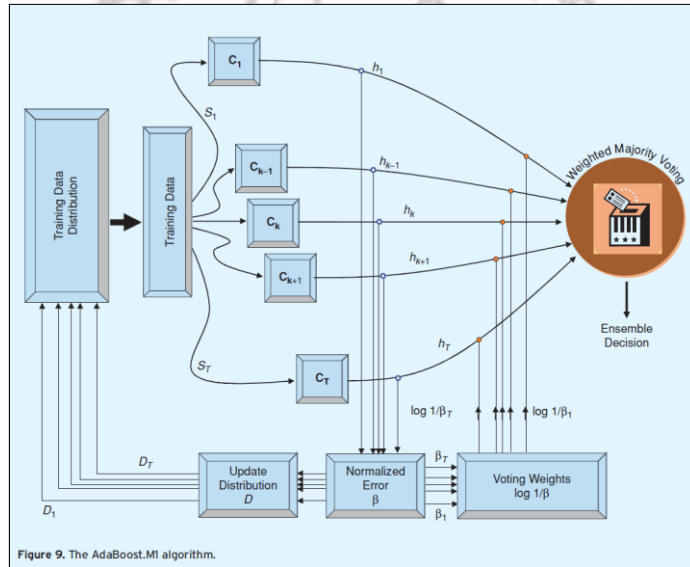


圖 2-4、Polikar 提出的其中一種決策架構 (Polikar, 2006)

在標註專利文本的決策方式時，會使用類似的架構，進行權重加總，以獲得各工程參數與專利文本之間的相關度分數。

## 第五節 RAG

在處理需要外部知識支援的任務上時，大語言模型雖然在語言理解上已達高準確度，但仍面臨知識更新困難與推理準確性的挑戰。Lewis et al. (2020) 提出檢索增強生成 (Retrieval-Augmented Generation, RAG) 架構，有效結合檢索系統與生成模型，在增強大語言模型文本生成的準確性與語意相關性下，Fan et al. (2024) 的研究中證明其架構可以減少幻覺的產生。圖 2-5 為 RAG 主要流程架構。

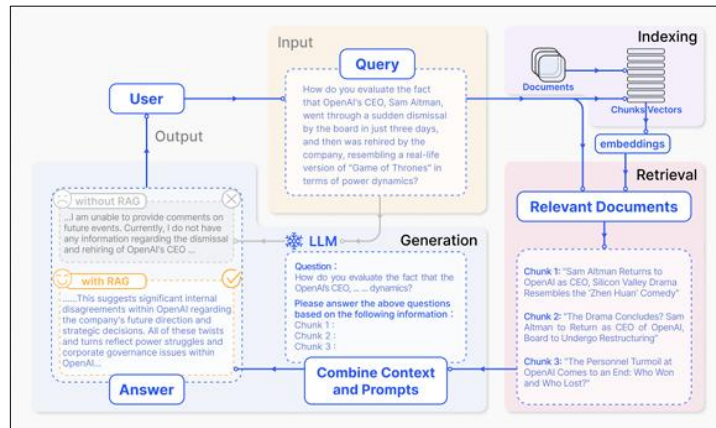


圖 2-5、RAG 架構流程(Gao et al. 2024)

RAG 主要可分為兩部分：檢索 (retrieval) 與生成 (generation)，當使用者進行查詢時，檢索器(retriever)會對外部資料庫進行最大內積檢索(Maximum Inner Product Search, MIPS) 尋找排名前  $K$  筆相關的資料，將檢索的內容與與查詢結合，輸入到大語言模型中，由生成器(generator)產生回答給使用者，使模型能即時更新知識而無需重新訓練，增加了可擴充性以及實用性，其詳細架構如圖 2-6 所示。

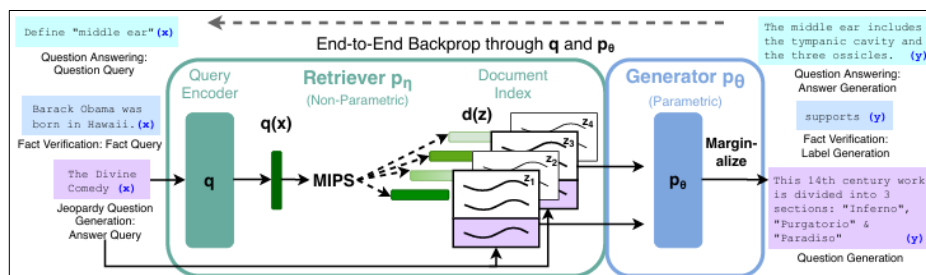


圖 2-6、RAG 詳細架構 (Lewis et al. 2020)

### 2.5.1 檢索器

由 RAG 的架構就可以知道其主要核心就是檢索器上的設計，檢索器的設計可以分為兩大類：稀疏檢索(sparse retrieval) 與密集檢索(dense retrieval)。稀疏檢索是基於詞彙的檢索方式，來自於比較傳統的自然語言處理技術，例如：TF-IDF 以及 BM25(Wang et al. 2023)，因為其高效率、高精確、可預測與低計算成本等特性，適合用於需要高精度的檢索任務，所以仍被廣泛使用。而其缺點是只能使用相似度檢索，無法處理複雜的需求，因此常與大語言模型搭配使用。

密集檢索又稱向量檢索(vector search)，如 Dense Passage Retriever (DPR) 與 RocketQA 透過語言模型將查詢與文件降至低維度的語意空間，能更好地抓取語意關聯，在開放式問答與語意檢索任務中展現更高的召回率與準確性 (Arabzadeh et al., 2021)，但 Chuang (2024) 的研究也指出，密集檢索雖然能在語意上表現更好，但可能會因資訊瓶頸而忽略細微詞彙的差異。

以上研究顯示，在設計檢索器時，選擇適當的檢索策略是相當重要的。對於需要高度精確詞彙匹配的任務，稀疏檢索比較有優勢；而在語意上的任務中，密集檢索又更勝一籌，搭配使用可以更有效的使用資源與提升檢索效率。

### 2.5.2 向量資料庫

在 RAG 架構中，向量資料庫 (vector database) 為整個系統運作的核心之一。其運作方式是將輸入文本與查詢內容轉換為向量後，利用語意相似度進行檢索。此方法不僅能降低原始文字儲存所需的空間，也因向量化後得以進行高效的數值計算，進而提升系統整體效能。然而，此技術也有限制，不同嵌入向量模型 (embedding vector model) 所產生的向量維度並不相同。由於維度大小會影響記憶體使用量與運算複雜度，若未處理維度不一致的情況，將可能對語意比對的準確性與檢索效率造成不利影響。

嵌入向量模型的向量維度由模型設計者預先定義，並通常於官方文件、原始論文或在 Hugging Face 等開源平台上公開標示。Hugging Face 為目前最具代表性的自然語言

處理開源平台之一，提供大量的預訓練模型、資料集與工具庫，使開發者能快速的建置、測試與部署語言模型相關應用。

本研究採用的 SBERT 模型即來自 Hugging Face，其原始論文明確指出向量維度為 384。相比之下，OpenAI 官方所提供的 text-embedding-ada-002 模型，其向量維度則高達 1536。雖然較高的向量維度有助於捕捉更細緻的語意特徵，但相對地也會提高計算成本。因此，嵌入模型的維度設計需在語意表現與計算效率之間取得平衡。

由於本研究的核心目標為於大量專利文本中提升語意檢索效率，SBERT 的中低維度設計使其成為兼具精度與效能的理想選擇。

### 2.5.3 LangChain 工具

前幾節只闡述了 RAG 架構的原理與概念，特別是向量資料庫進行語意檢索的邏輯，但在實作層面上，包括如何將文本透過嵌入模型轉換為向量、如何建置向量資料庫，以及如何整合上述架構以形成具備實際應用能力的完整系統，沒有進一步的說明。

LangChain 為一款由 Chase (2022) 所推出的開源應用框架，設計目的是協助開發者可以快速整合大型語言模型與各種工具，進行具備檢索能力與推理機制的語言模型應用開發。該框架支援 Python 語言與 Hugging Face 的生態，並廣泛應用於自然語言處理相關研究。

根據 LangChain 官方網站所整理的文獻列表，近年已有多篇學術論文採用 LangChain 作為研究與實驗工具。其中 Rackauckes (2024) 與 Jeong et al. (2024) 的研究皆針對 RAG 架構進行探討，並以 LangChain 作為主要開發與實驗工具。由於本研究著重於語意檢索的實作，故選用 LangChain 作為嵌入模型整合、向量資料庫的建置與檢索流程設計的應用框架。

# 第六節 專利探勘

在眾多專利資料庫中，使用美國專利資料庫(USPTO) 作為研究用的資料來源。因為其資料庫資訊的完整性高，不僅涵蓋了自美國 1790 年的專利資料，1976 年後的專利文本提供了全文檢索，可以獲取完整的專利資訊，圖 2-7 為其中一份專利文本的部分樣式。

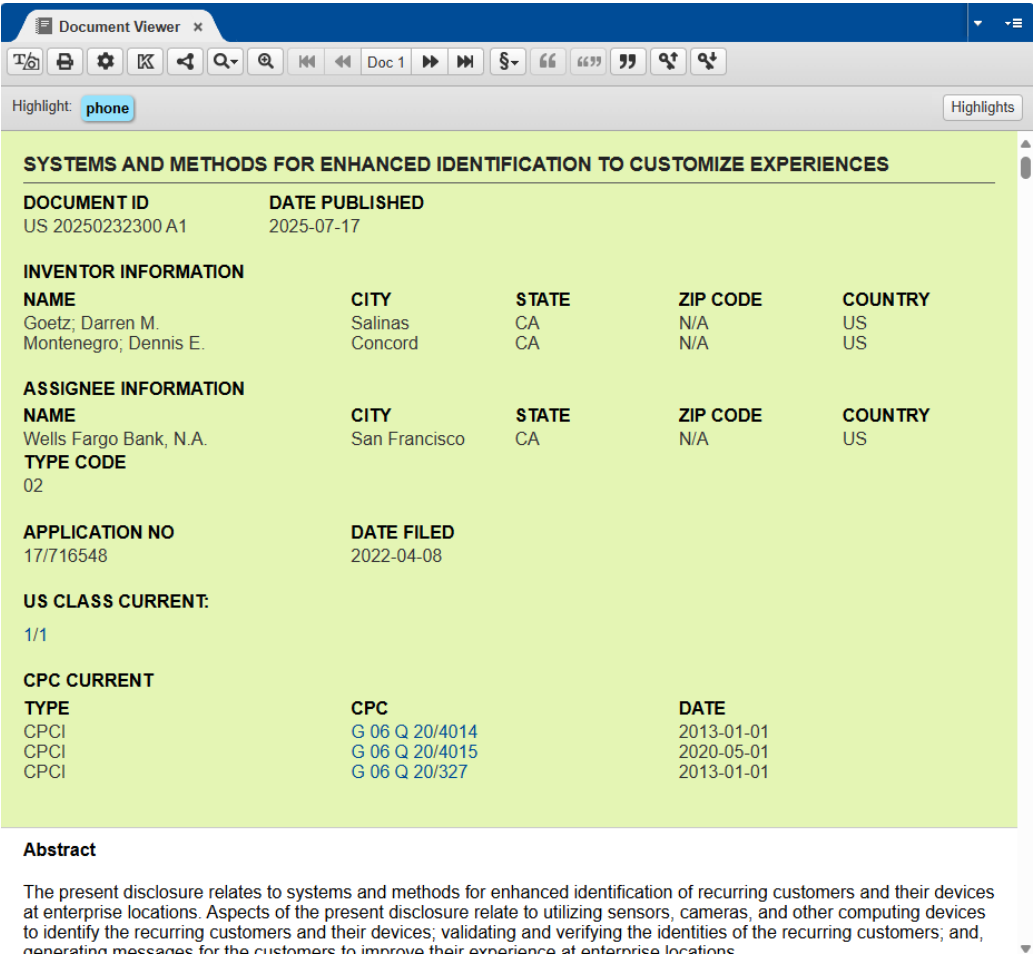


圖 2-7、USPTO 專利文本資訊

可以看到有相當完整的專利描述與資訊，其中還有描述專利的資訊欄位有：摘要 (abstract)、背景/總結(background/summary)、描述(description)、請求項(claims)。我們會取用這些段落，進行文本分析。另外，會取用「請求項」是因為其內容為整份專利文件最為關鍵的地方，定義了申請人的法律保護技術範圍，會明確指出發明人的專利技術內容，因此是必要分析的文本內容。



### 第三章 研究方法

本論文的研究方法流程如圖 3-1，首先，從美國專利資料庫(USPTO) 提取專利文本內容含專利號碼、標題、摘要、描述、請求項。由於專利文本的內容皆為原始資料，可能含有特數字元，所以需要做預處理，將特殊字元移除。接著，定義各項工程參數的關鍵字詞庫，增強對每項工程參數的意義。

工程參數資料以及關鍵字詞庫準備好後，使用語意分析(semantic)、TF-IDF、關鍵字匹配三種方式分析各段落的文本，也給予各段落權重，表示各段落的重要性。經過分數加總並篩選後，將有符合分數的工程參數標記於專利文本中，表示該項工程參數與其專利文本的相關度，並與原本的資料文本整合成新的資料表，因此每篇專利文本都有被標記的工程參數以及對應的相關分數。

最後使用 RAG 架構中的檢索器，讀取使用者的輸入字詞，進行轉化後，檢索整合有標記的專利文本資料，推薦出 10 篇最相關的專利文本並呈獻給使用者。

本次研究使用設備、套件與軟體如下：

#### 1. 電腦硬體設備：

- 作業系統：Windows 11
- 處理器：Intel(R) Core(TM) i9-9900K CPU @ 3.60GHz
- 記憶體：32 GB
- 顯示卡：Nvidia GeForce RTX 3060 Ti
- 顯示卡記憶體：24 GB

#### 2. 開發環境：

- Python 版本：3.12
- 使用套件：PyTorch、CUDA 11.8、Polars、Sentence-Transformers、Langchain-HuggingFace、sklearn
- IDE：Visual Studio Code、DataSpell

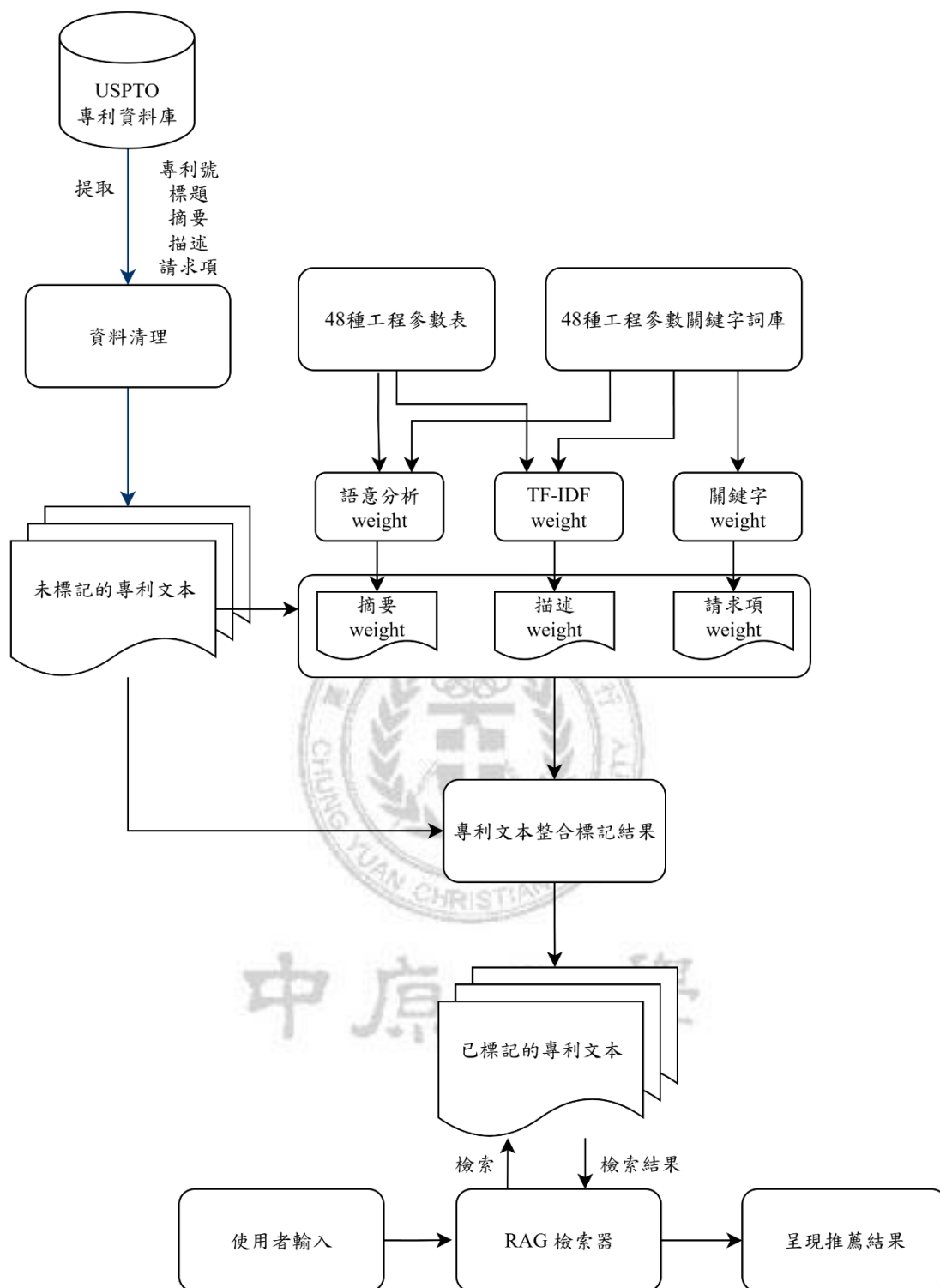


圖 3-1、研究方法流程圖

## 第一節 專利文本處理

本研究室已有事先從美國專利資料庫收集 2012 年至 2016 年六年的文本資料，共 98 萬篇專利文本，而在收集這些文本資料時，將原本專利文本的「背景/總結」欄位與「描述」欄位合併為 description 欄位，以利後續研究。為了進行本實驗，只從中隨機抽取 5000 筆資料進行標籤與分析。

首先，檢視原始資料查看資料欄位，原始資料如圖 3-2 所示。

app_...	invention_title	abstract	description	claims
13538195	"Programmable low pow...	"Disclosed is a meth...	"BACKGROUND 1. Field of...	"1. A multi-modulus divider (MMD) comprising: a c...
13248165	"Hearing aid device f...	"With a hearing aid ...	"CROSS-REFERENCE TO RELAT...	"1. A method for frequency compression of an input...
12569987	"Battery cooling appa...	"A battery pack appa...	"BACKGROUND The present...	"1. A battery pack apparatus comprising: a plurali...
13656105	"Methods of inhibitin...	"This invention offe...	"BACKGROUND OF THE INVENT...	"1. A method of inhibiting the growth of cyanobact...
14104237	"Method for welding g...	"Relating to electro...	"CROSS-REFERENCE TO RELAT...	"1. A method for welding a gold-silicon eutectic c...
14045743	"Vehicle roof support...	"A roof rail assembl...	"TECHNICAL FIELD The pr...	"1. A roof rail assembly comprising: an outer bod...
13322059	"Sandwich structured ...	"The invention relat...	"RELATED APPLICATIONS T...	"1. A structural member comprising: a first and a...
14550813	"Two stroke, opposed-...	"In a two-stroke opp...	"PRIORITY This applicat...	"1. A method of operating a two-stroke, opposed-pi...
13208884	"Process for producin...	"A medical resin com...	"CROSS-REFERENCE TO RELAT...	"1. A process for producing a molded article suita...
13838307	"Numerically controll...	"A coffee grinder us...	"FIELD OF THE INVENTION	"1. A method of grinding coffee comprising the ste...
14558222	"Noise shielding tech...	"A device having an ...	"CROSS REFERENCE TO OTHER...	"1. A device having an integrated well, comprising...
13050598	"Valve train with cam...	"A valve train for a...	"CROSS REFERENCE TO RELAT...	"1. A valve train for an internal combustion engin...
13625679	"Generating time slot...	"A time slot of regu...	"BACKGROUND In some ind...	"1. A computer implemented method to plan across t...
13415653	"Intelligent patch ch...	"A client computer s...	"RELATED APPLICATIONS T...	"1. A method, including: storing on a server syst...

圖 3-2、提取欄位後的資料

再檢視其中的內容，發現 description 欄位的內容，有特殊字元存在以及許多空格，如圖 3-3 所示。

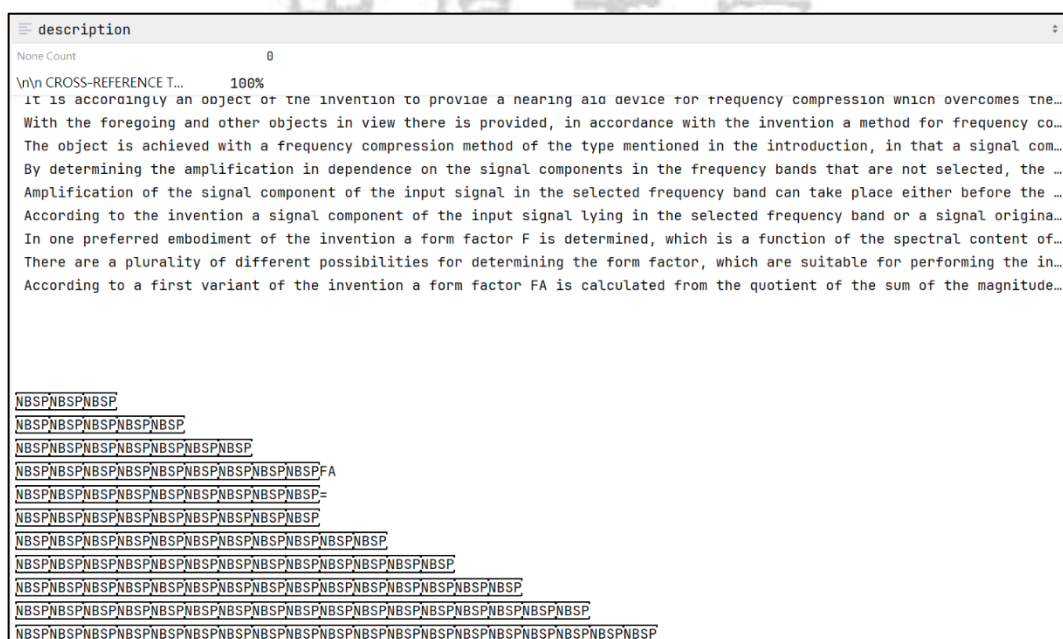


圖 3-3、原始文本內容

接著，清理特殊字元與多餘的空格，圖 3-4 為進行資料處理後的資料，圖 3-5 為所有資料進行處理後的概觀。(由於篇幅關係，只擷取部分內容)

```
"description": "CROSS-REFERENCE TO RELATED APPLICATION This application claims the priority, under 35 U.S.C. §119(e), of provisional
application No. 61/387,519, filed Sep. 29, 2010; the prior application is herewith incorporated by reference in its entirety. BACKGROUND OF
THE INVENTION Field of the Invention The invention relates to a method for frequency compression of an input signal entering a hearing aid
device and includes the below described steps. A frequency range that can be transmitted by the hearing aid device can be split into a
number of frequency bands. A number of adjacent frequency bands are combined into at least one group of frequency bands. One of the
frequency bands is selected from the group of frequency bands and the selected frequency band is moved into a target frequency band. The
invention also relates to a hearing aid device for performing such a method. To compensate for a user's individual hearing loss,
frequency-dependent amplification of an input signal entering the hearing aid device normally takes place in the hearing aid device. The
dynamic range, i.e. the range between the hearing threshold and the discomfort threshold, is generally very limited for a hearing-impaired
person compared with a person with normal hearing. Therefore modern hearing aid devices generally also perform dynamic compression by an
automatic amplification or gain control AGC. However there are hearing losses which cannot be compensated for in a satisfactory manner by
simple frequency-dependent amplification of an acoustic input signal. Examples of this are hearing losses with dead frequency ranges, in
which spectral components of the acoustic input signal cannot be made audible even with a high level of amplification. German patent
application DE 10 2006 019 728 A1 discloses a method for adjusting a hearing aid apparatus, in which a portion of an input signal spectrum
at a first frequency is amplified and shifted to a second frequency as a function of time, in order on the one hand to achieve a high level
of spontaneous acceptance of the hearing system due to an almost undistorted sound pattern of the hearing system between two adaptation
steps and on the other hand to assist the learning and acclimatization process of the hearing-impaired person in respect of the new
frequency pattern. One possibility for resolving the last-mentioned problem is what is known as frequency compression. Here spectral
components within a source spectral range (typically at higher frequencies) are shifted to a target frequency range (typically at lower
frequencies). Unlike the signal components of the acoustic input signal in the source frequency range, the signal components shifted to the
target frequency range can be made audible by amplification in this frequency range. One possible frequency compression method contains
the now described method steps. A frequency range that can be transmitted by the hearing aid device is split into a number of frequency
bands (channels). A number of adjacent frequency bands are combined into at least one group of frequency bands. One of the frequency bands
is selected from the group of frequency bands. The selected frequency band is moved to a target frequency band. Optionally the frequency
bands from the group of frequency bands that are not selected are suppressed. One disadvantage of this procedure is however that sound
signals which mainly differ in the content of the channels not selected are difficult to distinguish or can no longer be distinguished
after frequency compression. SUMMARY OF THE INVENTION It is accordingly an object of the invention to provide a hearing aid device for
frequency compression which overcomes the above-mentioned disadvantages of the prior art devices of this general type, which makes it
easier to distinguish sound signals which mainly differ in the content of the channels not selected. With the foregoing and other objects
in view there is provided, in accordance with the invention a method for frequency compression of an input signal entering a hearing aid
```

圖 3-4、進行處理後的文本內容

doc_num	title	abstract	description	claims
13538195	"Programmable low power mul...	"Disclosed is a method and ...	"BACKGROUND 1. Field of th...	"1. A multi-modulus divider (MMD) compris...
13248165	"Hearing aid device for fre...	"With a hearing aid device ...	"CROSS-REFERENCE TO RELATE...	"1. A method for frequency compression of-
12569987	"Battery cooling apparatus ...	"A battery pack apparatus h...	"BACKGROUND The present di...	"1. A battery pack apparatus comprising a...
13656105	"Methods of inhibiting cyan...	"This invention offers an e...	"BACKGROUND OF THE INVENTI...	"1. A method of inhibiting the growth of ...
14104237	"Method for welding gold-si...	"Relating to electronic com...	"CROSS-REFERENCE TO RELATE...	"1. A method for welding a gold-silicon e...
14045743	"Vehicle roof support struc...	"A roof rail assembly for a...	"TECHNICAL FIELD The prese...	"1. A roof rail assembly comprising: an o...
13322059	"Sandwich structured constr...	"The invention relates to a...	"RELATED APPLICATIONS This...	"1. A structural member comprising: a fir...
14550813	"Two stroke, opposed-piston...	"In a two-stroke opposed-pi...	"PRIORITY This application...	"1. A method of operating a two-stroke, o...
13208884	"Process for producing resi...	"A medical resin compositio...	"CROSS-REFERENCE TO RELATE...	"1. A process for producing a molded arti...
13838307	"Numerically controlled esp...	"A coffee grinder using a s...	"FIELD OF THE INVENTION Th...	"1. A method of grinding coffee comprisin...
14558222	"Noise shielding techniques...	"A device having an integra...	"CROSS REFERENCE TO OTHER ...	"1. A device having an integrated well, c...
13950598	"Valve train with camshaft ...	"A valve train for an inter...	"CROSS REFERENCE TO RELATE...	"1. A valve train for an internal combust...
13625679	"Generating time slot for m...	"A time slot of regular tim...	"BACKGROUND In some indust...	"1. A computer implemented method to plan...
13415653	"Intelligent patch checker"	"A client computer sends ap...	"RELATED APPLICATIONS The ...	"1. A method, including: storing on a ser...
13001165	"Rotating prism scanning de...	"The present invention rela...	"CROSS-REFERENCE TO RELATE...	"1. A scanning device comprising a first ...
12597313	"Aqueous silicone emulsion ...	"Aqueous silicone emulsions...	"RELATED APPLICATIONS This...	"1. An aqueous silicone emulsion composit...

圖 3-5、處理後資料概觀

為了更貼合 Python 語言的使用，會將整理後的資料儲存成 json 檔案進行實驗。

## 第二節 48 種工程參數字詞庫建立

目前沒有任何 48 種工程參數的關鍵字詞的公開資料，而產品開發者在查詢時，會使用科學術語進行查詢。因此，使用了 chatGPT 根據 48 種工程參數的內容以及常見的科學用詞，建立關鍵字詞庫，部分字詞庫如圖 3-6 所示，完整關鍵字詞庫可見附錄二。

parameter_id	keywords
1	["weight", "mass", "heavy", "light", "moving", "mobile", "dynamic"]
2	["weight", "mass", "heavy", "light", "stationary", "fixed", "static"]
3	["length", "angle", "long", "short", "dimension", "angular", "moving", "mobile"]
4	["length", "angle", "long", "short", "dimension", "angular", "stationary", "fixed"]
5	["area", "surface", "coverage", "moving", "mobile"]
6	["area", "surface", "coverage", "stationary", "fixed"]
7	["volume", "capacity", "size", "moving", "mobile"]
8	["volume", "capacity", "size", "stationary", "fixed"]
9	["shape", "form", "geometry", "configuration", "design", "contour"]
10	["amount", "substance", "quantity", "material", "matter", "mass"]
11	["amount", "information", "data", "knowledge", "content", "details"]
12	["duration", "time", "period", "lifetime", "moving", "action"]
13	["duration", "time", "period", "lifetime", "stationary", "action"]
14	["speed", "velocity", "fast", "slow", "rate", "acceleration"]
15	["force", "torque", "push", "pull", "tension", "compression", "twist"]
16	["energy", "power", "consumption", "efficiency", "moving", "dynamic"]
17	["energy", "power", "consumption", "efficiency", "stationary", "static"]
18	["power", "energy", "electrical", "mechanical", "output", "wattage"]
19	["stress", "pressure", "strain", "load", "mechanical", "compression"]
20	["strength", "durability", "robust", "strong", "weak", "mechanical"]

圖 3-6、48 種工程參數關鍵字詞庫(部分)

## 第三節 標籤專利文本流程

將文本資料進行清理以及建立好關鍵字詞庫之後，就可以開始對專利文本進行標籤的實驗，首先，定義好三種文本分析的分類器權重，本研究根據文獻探討，使用詞彙頻率與語意分析的組合，最終採用三種方式對文本進行分類：關鍵字匹配、TF-IDF 與語意分析，分別的權重是：0.2、0.3、0.5。

關鍵字匹配與 TF-IDF 分類方式是以詞彙頻率進行計算，語意分析可解釋專利文本中與工程參數及其關鍵字詞的語意，而本研究期望結果可泛用化，因此兩種方式的權重各給了 0.5，而詞彙計算中，TF-IDF 是以前後文進行詞彙計算，其結果可解釋詞彙的相關性，因此權重給予 0.3，關鍵字匹配則給予 0.2 的權重。

### 3.3.1 各分類器分數計算方式

關鍵字匹配的計算公式如下：

- $T$  為輸入的原始文本
- $k$  為關鍵字
- 函數  $C(k, T)$  代表關鍵字  $k$  在文本  $T$  中出現的次數

$$C(k, T) = \sum_{i=1}^n \delta(w_i = k) \quad \text{公式(4)}$$

$\delta(w_i = k)$  如果為相符的單字為 1，否則為 0。這樣就可以得到單一工程參數中的關鍵字詞在該文本的關係分數。

TF-IDF 分類器的流程則是會對文本與工程參數進行 TF-IDF 的計算，將計算結果的向量再進行餘弦相似度計算，得到工程參數與文本之間的相關分數。

語意分析則是使用 SBERT 計算，分別計算工程參數對文本與關鍵字詞庫對文本的分數進行加總，獲得工程參數對文本的關係分數。

### 3.3.2 工程參數分數計算流程

當單一分類器計算所有工程參數對文本的分數後，對此分數進行標準化再乘上權重後，即可得到該分類器的每項工程參數對該文本的相關分數，公式如下：

- $S = \{s_1, s_2, \dots, s_n\}$ ， $S$  為分類器， $s_n$  為工程參數  $n$  的分數
- $s_{max} = \begin{cases} \max(S), & \text{if } \max(S) > 0 \\ 1, & \text{if } \max(S) = 0 \end{cases}$
- $w$  為權重

$$s'_i = \frac{s_i}{s_{max}} \times w \quad \text{公式(5)}$$

再將各分類器的每項工程參數所得到的分數進行加總，就會得到該工程參數對文本的總分。



### 3.3.3 專利文本標記方法

在前述架構中，我們已說明各工程參數在文本中的計算方式。然而，專利文本中的不同段落對整體專利的重要性並不相同。根據文獻探討中的決策架構，本研究對「摘要」、「描述」與「請求項」分別賦予不同權重，分別為 0.2、0.3、0.5。

「請求項」作為專利的核心與技術保護範圍，具有最高的重要性，因此給予 0.5 的權重；「描述」包含詳細技術細節與背景，支撐整體技術邏輯與可實現性，因此給予次高的 0.3；「摘要」則提供概要性的說明，相對而言重要性較低，故設為 0.2。

根據上述權重架構，可將每段落文本分別輸入各分類器進行工程參數分析，獲得該參數對每段落的分數。再依照各段落的權重進行加權總和，即可得出該工程參數對整篇專利文本的整體相關性分數，如圖 3-7 所示。

patent_num	title	params_score
13538195	"Programmable low power multi-modulus divider with 50/...	{0.066835,0.012696,0.162295,0.122192,0.094992,0.058057...
13248165	"Hearing aid device for frequency compression"	{0.01061,0.206636,0.031799,0.106627,0.218272,0.28674,-...
12569987	"Battery cooling apparatus for electric vehicle"	{0.04476,-0.068535,0.155326,0.041257,0.410133,0.337608...
13656105	"Methods of inhibiting cyanobacteria growth by adminis...	{0.138722,0.154259,0.135657,0.154025,-0.003579,0.05695...
14104237	"Method for welding gold-silicon eutectic chip, and tr...	{0.011797,0.034729,0.136263,0.159122,0.062373,0.120204...
14045743	"Vehicle roof support structure including side curtain...	{0.006165,0.068761,0.04835,0.144469,0.846994,0.98425,0...
13322059	"Sandwich structured construction element"	{0.103609,0.154047,0.346153,0.423207,0.237112,0.293263...
14550813	"Two stroke, opposed-piston engine with engine braking"	{0.085714,0.17647,0.197273,0.233233,0.135921,0.217886,...
13208884	"Process for producing resin composition and process f...	{-0.082587,0.028211,0.040108,0.193569,-0.05703,0.14253...
13838307	"Numerically controlled espresso grinder"	{-0.142969,0.079278,-0.01472,0.209237,-0.108905,0.0631...
14558222	"Noise shielding techniques for ultra low current meas...	{0.019787,-0.033721,0.089279,0.022974,0.160317,0.12072...
13050598	"Valve train with camshaft with an axially displaceabl...	{0.087956,0.149827,0.201268,0.238789,0.083972,0.134527...
13625679	"Generating time slot for managing time across multipl...	{0.028912,0.050022,0.176117,0.183151,0.098883,0.119951...
13415653	"Intelligent patch checker"	{0.03684,0.085782,0.166714,0.244731,0.111203,0.180225,...
13001165	"Rotating prism scanning device and method for scannin...	{0.165295,0.200213,0.40476,0.504847,0.449005,0.489045,...
12597313	"Aqueous silicone emulsion for imparting water repelle...	{0.087245,0.028864,0.080725,0.020375,0.26491,0.230902,...
13624493	"PEGylated mutated clostridium botulinum toxin"	{0.053204,0.140087,0.032867,0.105391,-0.03249,0.062888...
13322596	"Structure for transporting, installing and dismantlin...	{0.105851,0.161071,0.166153,0.187333,0.116755,0.107682...
12066735	"Coordinate position detecting device"	{-0.002878,0.092065,0.180802,0.207825,0.176743,0.19418...
12116018	"Packet structure for a mobile display digital interfa...	{0.192623,0.120286,0.3388,0.193928,0.254009,0.098121,0...

圖 3-7、專利文本工程參數分數表

最終，建立了每篇專利文本對各項工程參數所對應的分數資料表。透過該表格進行資料庫檢索時，篩選分數較高的工程參數進行比對，而不需比對整份文本，從而提升整體檢索的效率。

## 第四節 應用檢索器

由於本研究以不調整模型為目標，因此在轉換使用者查詢語句及專利文本上，使用 SBERT 嵌入模型，向量資料庫則使用 LangChain 所提供的 Chroma 向量資料庫。

在 LangChain 官方教程中，應用 RAG 架構需要進行兩步驟，第一步是將文本向量索引化儲存至向量資料庫，第二步是檢索與產生回答，但本研究著重於加快檢索效率，因此不會進行產生回答的步驟，接下來逐一說明應用 RAG 的流程。

### 3.4.1 向量資料庫索引化

詳細索引化(indexing)的步驟如圖 3-8 所示。

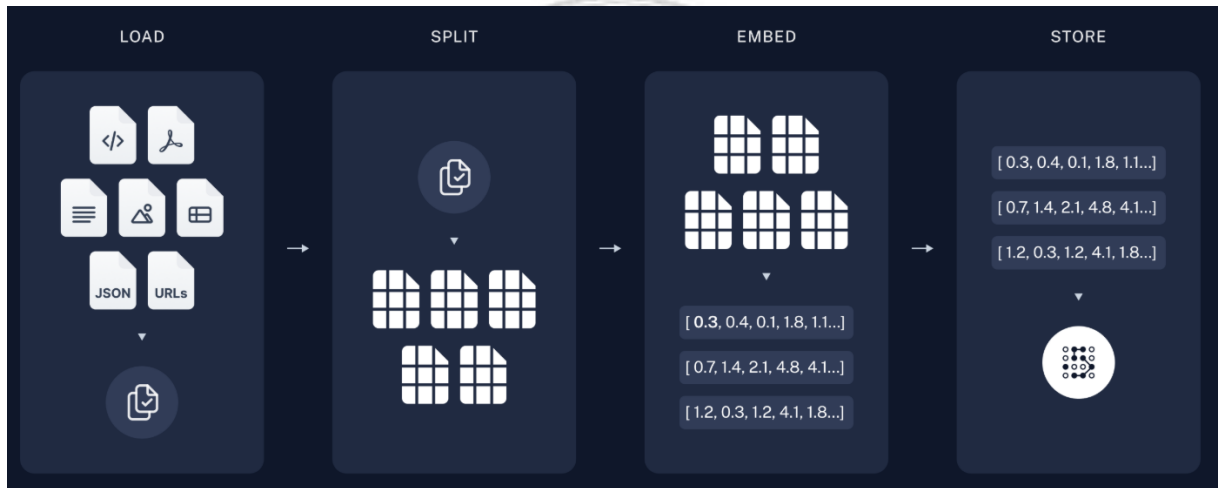


圖 3-8、索引化步驟 (Chase)

由於程式中的參數與變數眾多，逐一說明會過於複雜，因此以虛擬碼說明程式邏輯與流程。首先，引入要使用的套件，載入整理好的專利文本與工程參數分數對照表，設定好嵌入模型與文本分割器的參數，如圖 3-9 說明。

```
Import the required libraries: json, langchain_huggingface, langchain_chroma,
langchain_core.documents, langchain.text_splitter
Load patent_data.json and patent_score_data.json and integrate to one dictionary named
“patents”
Initialize textsplitter object with chunk size of 1000, chunk overlap of 200, set track_index
True
Initialize embedding_model object with model name “all-MiniLM-L6-v2”, using cuda
device and set normalize embeddings True
```

圖 3-9、檢索器虛擬碼說明



在這邊文本分割器的參數參照官方文件的推薦，設定文本分割大小為 1000，保留上下文長度為 200 並啟用追蹤分割文本在原始文本中的位置功能。all-MiniLM-L6-v2 為 SBERT 嵌入模型的名稱，使用 GPU 比對文本，並啟用嵌入向量的標準化功能。

接著要篩選專利文本的內容，轉成 Document 物件，以符合 Chorma 資料庫的格式。首先只會加入相關分數高於 0.4 的工程參數，將這些參數整合成一字串，並使用文本分割器進行分割，將被分割的文本轉成 Document 物件並加到 documents 清單中，虛擬碼說明如圖 3-10。

```
Create an empty list named "documents" for storing document object
Perform a for loop using patent as variable traverse in patents:
    Create an empty list named "triz_param_names"
    For each patent's engineering parameters:
        If engineering parameter score is greater than 0.4:
            Add to triz_param_names
    Create a variable named "contents" by passing triz_parameters
    Use textsplitter to chunk contents and named as "chunked_text_set"
    For each chunk_text in chunked_text_set:
        Concat all triz parameters in triz_param_names named "triz_parameters"
        Concat "chunked_text" and "triz_parameters" as "contents"
        Initialize document object as page_content with "contents", metadata as "patents"
        Append document object to "documents"
```

圖 3-10、檢索器虛擬碼說明

### 3.4.2 檢索向量資料庫

將向量資料庫準備好後，就可以使用查詢語句進行向量資料庫檢索，其虛擬碼如圖 3-11 所示。

最後，將整合好的專利文本使用 SBERT 模型建置到 Chroma 向量資料庫內轉成向量維度，設定檢索器搜索參數為相似度，尋找 10 篇相關度最高的專利文本。將使用者的查詢語句也透過 SBERT 模型轉成同個維度的向量在 Chroma 向量資料庫進行文本相關度計算，並將最相關的 10 篇專利文本及其分數記錄下來，以利後續分析。

Create an object named “vectorstore” by passing Chorma.from\_documnets object as documents with “documents” and embedding with “embedding\_model”

Create an object named “retriever” by “vectorstores.as\_retriever” as search type with “similarity” and search\_kwarg with a dictionary key is “k”, its value is 10

Create a String object named “query” by passing a query String

Create an object named “related\_patents\_with\_scores” by “vectorstore.similarity\_search\_with\_scores” with query

圖 3-10、檢索器虛擬碼說明

## 第五節 分析結果流程

由於單篇專利文本通常包含約一萬至三萬字，難以用人工方式進行結果驗證，因此本研究設計六組查詢語句，並以兩種版本的專利資料進行實驗比較。一為原始全文，另一為僅保留已標記工程參數的內文。透過比對查詢語句與工程參數的匹配程度及內文相關性，進行數據分析，並針對推薦的十篇專利文本中工程參數的出現頻率進行統計，以評估其檢索能力。

查詢語句的設計依據工程、技術、需求、TRIZ、陳述與目標導向分析(Chen et al. 2024)，於是設計出以下問題，如表 3-1 所示：

表 3-1、查詢語句表

Type	Query
Engineering	How can mechanical friction be reduced to improve efficiency?
Technical	What methods can prevent thermal deformation during plastic molding?
Requirement	What technologies can reduce device size while maintaining functionality?
TRIZ	What solutions simultaneously enhance system stability and response speed?
Statement	An electric motor produces excessive vibration, causing fatigue failure.
Goal	How can solar module absorption efficiency be improved?

## 第四章 研究結果

本章節將針對前述實驗結果進行深入討論與分析，內容分為兩個部分：第一部分著重於專利文本標籤結果之探討，第二部分則分析六組語句透過檢索器所獲取資料的相關結果。

### 第一節 專利文本標籤的結果

由於專利文本數量龐大且內容繁複，逐一進行分析不僅耗時也不具效率，故本研究採隨機抽樣方式，選取其中五篇專利文本，並針對其標籤結果進行探討，進一步分析各文本中排名前五之工程參數，以進行具體比較與觀察。

這五篇的專利編號與標題如下表所示。

表 4-1、五篇專利文本編號與標題

專利編號	標題
14539847	Electronic system with data management mechanism and method of operation thereof
12385303	Image processing apparatus and image processing method, and recording medium for processing breast image based on local contrast values in a local region in a mammary gland region of breast image
13794125	Spatially coherent nearest neighbor fields
12397463	Image forming apparatus and paper-jam clearing method
12393670	Method and apparatus for channel encoding and decoding in a communication system using low-density-parity-check codes

我們列出專利編號 14539847 的前五名分數、摘要以及個段落分數最高的工程參數，進一步分析了解與工程參數之間的關係，如表 3-3。

表 4-2、專利編號 14539847 分析表

"An apparatus includes: a media; a head over the media; a head actuation motor coupled to the head; control circuitry, coupled to the head actuation motor, configured to position the head; and an environmental sensor, coupled to the control circuitry, configured to measure an environmental condition; wherein the control circuitry is further configured to: perform a
---

background task; and adjust a frequency of the background task based on the magnitude of the environmental condition."		
前五名工程參數		相關分數
31. "Other harmful effects generated by system"		0.508876
40. "Other Harmful Effects Acting On the System"		0.366419
33. "Compatibility or connectability"		0.358772
43. "Automation"		0.313001
34. "Ease of operation"		0.298507
段落	最高分工程參數編號	分數
摘要(Abstract)	43. "Automation"	0.74152
描述(Description)	31. "Other harmful effects generated by system"	0.17125
請求項(Claims)	33. "Compatibility or connectability"	0.10174

儘管摘要內容未明確提及「自動化」相關詞彙，僅出現如驅動馬達、電路控制、感測元件以及背景中執行任務等描述，我們的分類器仍能準確判定其與「自動化」高度相關，展現了其對語句隱含意義的推理能力。另一方面，「有害的副作用」在工程參數評分中名列前茅，可能是因為文本其他段落中存在與其高度關聯的內容，顯示出在文本不同段落上給予權重，有達到預期的效果。

接著看專利編號 12385303 的相關分數表，如表 3-4 所示。

表 4-3、專利編號 12385303 分析表

" An image processing apparatus includes: an image obtaining device which obtains a breast image obtained by radiography of a breast; a mammary gland region extracting device which extracts a mammary gland region from the breast image; a local region setting device which sets a plurality of local regions around pixels belonging to the extracted mammary gland region; a local contrast value calculating device which calculates a plurality of local contrast values in a local regions, for each of the set plurality of local regions; and an image processing device which applies image processing to the breast image on the basis of the calculated plurality of local contrast values. Thus, considering a contrast between a mammary gland and
--

a fat region, a stable image processing result can be obtained while enhancing viewability of a local mammary gland structure and a lesion.."		
前五名工程參數		相關分數
33. "Compatibility or connectability"		0.356989
31. "Other harmful effects generated by system "		0.319913
45. "System complexity"		0.298760
37. "Security"		0.278518
13. "Duration of action of stationary object "		0.274580
段落	最高分工程參數編號	分數
摘要(Abstract)	45. "Ability to detect or measure"	0.730851
描述(Description)	33. "Compatibility or connectability"	0.5
請求項(Claims)	42. "Manufacture precision or consistency"	0.5

從這張表格，可以看出高分工程參數的多樣化：整體分數最高的參數不完全相同於各段落內最高分的參數，顯示不同文字段落突中，描述不同的技術焦點。

另外，前五名的工程參數與該專利的影像精度、系統複雜性、安全性等核心問題息息相關，有助於理解此篇專利文本考項在設計時的技術矛盾。

該專利的技術焦點似乎圍繞著「提升局部結構的可視性」與「提高影像處理穩定性」，而這正好對應到 TRIZ 中的工程參數的測量能力、連結程度與加工精度等參數。

表 4-4、專利編號 13794125 分析表

"In embodiments of spatially coherent nearest neighbor fields, initial matching patches of a nearest neighbor field can be determined at image grid locations of a first digital image and a second digital image. Spatial coherency can be enforced for each matching patch in the second digital image with reference to respective matching patches in the first digital image based on motion data of neighboring matching patches. A multi-resolution iterative process can then update each spatially coherent matching patch based on overlapping grid regions of the matching patches that are evaluated for matching regions of the first and second digital images. An optimal, spatially coherent matching patch can be selected for each of the image grid locations of the first and second digital images based on iterative interaction to enforce
---

the spatial coherency of each matching patch and the multi-resolution iterative process to update each spatially coherent matching patch."

前五名工程參數		相關分數
6. "Area of stationary object"		0.517035
5. "Area of moving object"		0.42672
36. "Repairability"		0.3976
11. "Amount of information"		0.360487
4. "Length or angle of stationary object"		0.347843
段落	最高分工程參數編號	分數
摘要(Abstract)	11. "Amount of information"	0.730851
描述(Description)	6. "Area of stationary object"	0.5
請求項(Claims)	11. "Amount of information"	0.5

在表 3-5 中，工程參數 Amount of information 在摘要與請求項中都是最高分項目，這暗示該技術的主要目標可能是最大化匹配區域提供的資訊量、提升圖像比對的準確性與解析度，但最高分是 Area of stationary object，可能表示這項技術用在動態的場景中又強調了靜態區域的幾何穩定性，另外 Repairability 分數排名第三，表示其影像在動態捕捉時可能有會有損壞需要修復影像的情形

表 4-5、專利編號 12397463 分析表

"A method and apparatus for encoding and decoding a channel in a communication system using a Low-Density Parity-Check (LDPC) code. The encoding method includes determining a modulation scheme for transmitting a symbol; determining a shortening pattern in consideration of the determined modulation scheme; grouping columns corresponding to an information word in a parity-check matrix of the LDPC code into a plurality of column groups; ordering the column groups; determining a range of a resulting information word desired to be obtained by shortening the information word; based on the range of the resulting information word, performing column group-by-column group shortening on the ordered column groups of the information word, according to the determined shortening pattern; and LDPC-encoding the shortened information word."



前五名工程參數		相關分數
34. "Ease of operation"		0.360331
11. "Amount of information"		0.343412
31. "Other harmful effects generated by system"		0.321447
33. "Compatibility or connectability"		0.314917
45. "System complexity"		0.268862
段落	最高分工程參數編號	分數
摘要(Abstract)	28. "Loss of information"	0.980935
描述(Description)	34. "Ease of operation"	0.185334
請求項(Claims)	33. "Compatibility or connectability "	0.119686

Ease of operation 在描述段落以及整篇專利中排名最高，意味著該技術設計操作流程有高度相關，在請求項中，Compatibility or connectability，排名很高，代表這項技術可能與相容性有關，而摘要有提到 LDPC 正是一種通訊架構。另外第三名的工程參數「其他有害影響」，可能在暗示其技術會造成精度或是變頻時導致系統惡化，

表 4-6、專利編號 12393670 分析表

"A paper conveying path extends from a paper feeding tray to an image forming unit, a branch conveying path extends from a branching point in the paper conveying path to a paper container, a first paper sensor and a second paper sensor are arranged in the paper conveying path upstream of the branching point, and a switch nail is arranged at the branching point to switch paths between the paper conveying path and the branch conveying path. The switch nail is controlled so that a jammed paper is conveyed to the paper container via the branch conveying path upon detecting paper jamming in the paper conveying path and the first sensor detects the jammed paper while the second sensor does not detect the jammed paper."		
前五名工程參數		相關分數
33. "Compatibility or connectability"		0.384119
31. "Other harmful effects generated by system"		0.376292
40. "Other Harmful Effects Acting On the System"		0.330081

11. "Amount of information"		0.276608
36. "Repairability"		0.273707
段落	最高分工程參數編號	分數
摘要(Abstract)	47. "Ability to detect or measure"	1.0
描述(Description)	33. "Compatibility or connectability"	0.5
請求項(Claims)	11. "Amount of information "	0.5

摘要提及卡紙偵測與紙張路徑切換機制，因此可推論「系統兼容性」是關鍵評估因素，理應優先考量。而「其他有害影響」得分次之，表示紙張傳送路徑可能導致故障率上升。為避免卡紙問題，系統設計中納入偵測機制，有助於提升資訊量降低系統故障率，也與請求項中得分最高的「資訊量」特性息息相關。

綜合以上分析結果可確認，透過本研究所構建的分類器與段落權重計算機制，確實能夠推論專利文本所要克服之技術問題，進而得到與某項工程參數的相關性。然而，其大部分工程參數的得分在整體相關性分數約介於 0.3 至 0.4 之間，顯示在關鍵字詞的選取及專利文本處理方法方面，尚有優化與提升的空間。

## 第二節 檢索器結果分析

上一節的主要在分析標籤的結果，確認是否對專利文本有解釋性，表示以方式此標記工程參數對檢索得結果是有幫助的。因此接著分析檢索器的給出的推薦結果並計算其檢索時間。

### 4.2.1 檢索效率驗證

首先，如表 3-8 所示，可驗證檢索所需時間。由結果可見，全文檢索在將文本載入所花時間為檢索工程參數的數倍以上。這過程需達十餘分鐘完成。這主要由於 SBERT 模型在處理長文本方面有所限制，需將單一專利文本切割為多個段落文本以進行相似度計算，此操作將導致需載入至向量資料庫的段落數量顯著增加，由原本的 5000 篇文本，增加至 255039 篇。



然而，在進行查詢語句檢索時，其效率並未因文本數量或字數差異而產生顯著落差，即使面對懸殊的資料規模，系統仍具備快速回應的能力。此結果反映出，目前檢索器於處理查詢語句方面具備一定效能，但在整體設計上仍有進一步優化的空間。

表 4-7、檢索花費時間比較表

檢索全文所花時間	檢索工程參數所花時間
Patent Search Results Query: How can mechanical friction be reduced to improve efficiency Total documents loaded: 255039 Setting retrieval model... End of retrieval. Retrieval model setting time: 694.30 seconds Start querying... End of query. Query time: 4.14 seconds	Patent Search Results Query: How can mechanical friction be reduced to improve efficiency Total documents loaded: 5000 Setting retrieval model... End of retrieval. Retrieval model setting time: 6.86 seconds Start querying... End of query. Query time: 4.21 seconds
Patent Search Results Query: What technologies can reduce device size while maintaining functionality? Total documents loaded: 255039 Setting retrieval model... End of retrieval. Retrieval model setting time: 789.32 seconds Start querying... End of query. Query time: 4.31 seconds	Patent Search Results Query: What technologies can reduce device size while maintaining functionality? Total documents loaded: 5000 Setting retrieval model... End of retrieval. Retrieval model setting time: 6.88 seconds Start querying... End of query. Query time: 4.23 seconds
Patent Search Results Query: What solutions simultaneously enhance system stability and response speed? Total documents loaded: 255039 Setting retrieval model... End of retrieval. Retrieval model setting time: 740.62 seconds Start querying... End of query. Query time: 4.24 seconds	Patent Search Results Query: What solutions simultaneously enhance system stability and response speed? Total documents loaded: 5000 Setting retrieval model... End of retrieval. Retrieval model setting time: 7.16 seconds Start querying... End of query. Query time: 4.27 seconds
Patent Search Results Query: An electric motor produces excessive vibration, causing fatigue failure. Total documents loaded: 255039 Setting retrieval model... End of retrieval. Retrieval model setting time: 656.61 seconds Start querying... End of query. Query time: 4.27 seconds	Patent Search Results Query: An electric motor produces excessive vibration, causing fatigue failure. Total documents loaded: 5000 Setting retrieval model... End of retrieval. Retrieval model setting time: 6.95 seconds Start querying... End of query. Query time: 4.24 seconds
Patent Search Results Query: How can solar module absorption efficiency be improved? Total documents loaded: 255039 Setting retrieval model... End of retrieval. Retrieval model setting time: 752.47 seconds Start querying... End of query. Query time: 4.24 seconds	Patent Search Results Query: How can solar module absorption efficiency be improved? Total documents loaded: 5000 Setting retrieval model... End of retrieval. Retrieval model setting time: 7.00 seconds Start querying... End of query. Query time: 4.16 seconds

## 4.2.2 檢索結果比較

表 4-8 是檢索全文與檢索工程參數的結果比較表。檢索工程參數的內容僅只有專利標題與工程參數，表中以紅字標示兩者共同包含的專利文本。經仔細分析後，可歸納出以下幾點值得探討的現象，並針對其可能原因進行說明。

表 4-8、檢索文本最高相關專利文本比較

查詢語句	檢索全文		檢索工程參數	
	專利編號	分數	專利編號	分數
How can mechanical friction be reduced to improve efficiency?	13702482	0.9268	12591370	1.1134
	14143802	0.9431	13403348	1.1392
	13702482	0.9637	13656576	1.1496
	13418161	0.967	13702482	1.1782
	13702482	0.9702	12532482	1.2346
	13702482	0.9906	12554287	1.2822
	13403348	0.9911	14006752	1.2884
	13418161	0.9977	13603900	1.3081
	13702482	0.9991	13333712	1.3369
	13702482	1.0033	13052947	1.3409
What methods can prevent thermal deformation during plastic molding?	10576148	0.5905	13657912	1.1012
	13612135	0.7372	10576148	1.1019
	10576148	0.8361	12665827	1.1485
	12534531	0.871	12667447	1.2085
	12534531	0.8717	12085276	1.224
	10576148	0.8725	13730775	1.2279
	14238292	0.878	11701789	1.2302
	10576148	0.8843	14553508	1.2712
	10576148	0.8903	14103523	1.29
	10576148	0.8914	13195653	1.3041
What technologies can reduce device size while maintaining functionality?	13742185	0.7486	14282541	1.0395
	13180301	0.8142	13768353	1.0622
	12333298	0.8375	14187089	1.1025
	14029623	0.8813	12587646	1.1428
	13831782	0.8846	12364396	1.1429
	12634114	0.9016	13872157	1.1508
	13782299	0.9029	11643167	1.1589

	13549719	0.9188	12070508	1.1589
	13176395	0.9228	13180301	1.162
	13590217	0.9291	12149486	1.1684
What solutions simultaneously enhance system stability and response speed?	10537566	1.117	13341501	0.9263
	12508814	1.1203	12508814	0.9815
	12894395	1.1245	12967587	1.1682
	13614467	1.1459	13139989	1.1864
	13510100	1.151	12466454	1.1879
	13030312	1.1535	11874258	1.1919
	12698328	1.1873	14010203	1.2053
	10537566	1.191	12125633	1.244
	13614467	1.1919	12433208	1.244
	10537566	1.1927	13235062	1.2445
An electric motor produces excessive vibration, causing fatigue failure.	13222257	0.8978	13222257	0.9093
	13007199	0.9112	13007199	1.0202
	12192059	0.9289	14124454	1.1029
	13222257	0.9558	12566292	1.1497
	13007199	0.9562	12303573	1.1565
	13222257	0.9632	12311515	1.1655
	14460210	0.9756	13763867	1.1791
	12192059	0.9819	14116041	1.188
	13007199	0.9848	13047950	1.1896
	12192059	0.9891	12192059	1.1908
How can solar module absorption efficiency be improved?	12579465	0.7703	13155914	1.0939
	13480393	0.8305	13621508	1.0979
	12914710	0.8629	14116624	1.1427
	12914710	0.8771	12982332	1.1636
	13480393	0.8912	12794307	1.165
	14116624	0.9004	12914710	1.1662
	12579465	0.9026	12579465	1.1976
	13480393	0.9125	12479136	1.2432
	13866315	0.9132	12623318	1.2449
	14116624	0.9133	12818712	1.2611

1. 全文檢索結果中出現重複的專利編號：此現象的原因是每篇切割的文本被當作獨立的文本去比對的關係，實驗中有嘗試要修正其輸出結果，但因為是透過 Chroma 套件中的比對函數所得出的推薦結果，其中沒有參數可以對此作調整，要使其沒有重複性的編號，需要自己設計檢索器或者增加輸出的最佳結果後再進行篩選，但不論使

用哪種方式，就不是在同樣條件下進行檢索比較，因此沒有對此部分進行修正。

2. 兩者結果均出現分數大於 1 的情形：通常出現此情形與分數計算公式或權重設定有關，例如文件長度、關鍵字密度或工程參數匹配度異常提升分數。然而，其主要原因可能是在語句查詢進行維度轉換的過程中，檢索器將該語句視為涉及多項工程參數的關係，顯示該檢索器具備解析多重目標語意的能力。
3. 推薦結果重疊率偏低：此現象反映出全文檢索與工程參數檢索所採用的語義模型或檢索邏輯差異顯著，代表兩者在判定文本相關性方面的依據與角度不盡相同。

由此表結果可推知，檢索器具備解析多重目標語意的能力，能夠辨識查詢語句中涉及的多項工程參數。然而，其具體比對與分數計算過程不可得，未能提供透明的推論機制。進一步觀察其推薦結果，可發現該檢索器在尚未進行調整的狀態下，其理解查詢語句與專利文本之間關聯的邏輯，與 TRIZ 所依循的問題分析與參數對應流程存在著顯著的差異。

#### 4.2.3 檢索結果文本分析

接著可以發現，第五與第六個語句查詢所對應的文本結果，其重疊率為所有查詢中最高。由於第五個查詢語句類型是問題描述，因此本研究選擇以此語句為分析基礎，以探討其在工程參數檢索上的表現。

此外，為了驗證僅透過工程參數進行檢索是否能提供對開發者具實用性的推薦結果，本段分析將聚焦於工程參數相關資料。為提升分析效率與可讀性，本文僅擷取每篇專利文本摘要的前 200 字進行比對與解釋。

表 4-9、檢索工程參數結果表

編號	標記工程參數	摘要
1	Other harmful effects generated by system	Disclosed is a horizontal vibration motor, the motor including a stator having a coil block secured to a bottom plate of a case, a vibrator having magnets placed above the coil block and a weight hold...

2	Noise, Ability to detect or measure Power, Energy used by stationary object, Energy used by moving object	An electric machine including a housing, a stator mounted within the housing, a rotor rotatably mounted within the housing relative to the stator, and a vibration sensor arranged within the housing. T...
3	Other harmful effects generated by system, Other Harmful Effects Acting On the System	A rotor for an electric machine includes a pole core having a coolable magnetisable rotor section which is made of a super-conducting material. The rotor section has a rotationally symmetric geometry....
4	Other harmful effects generated by system	A pump system. The pump system includes a pump, a motor coupled to the pump, a sensor coupled to a power source, and a controller. The motor operates the pump. The sensor detects an electrical charact...
5	Power, Amount of substance, Energy used by stationary object Energy used by moving object Other harmful effects generated by system, Strength	A rotor ( 11 ) for an electric machine with low-inertia permanent magnets ( 13 ) located between the poles ( 15 ) of the electric machine and a rotor hub, where the rotor hub is comprised of an intern...
6	Noise, Stress or pressure, Other harmful effects generated by system, Force or torque	A vibration isolator is described for inhibiting transfer of vibration to an apparatus during evacuation thereof by a pump. The isolator comprising a bellows for isolating from the ambient atmosphere ...
7	Other harmful effects generated by system, Other Harmful Effects Acting On the System	A vibrating actuator assembly including: a rotor that is rotatable; a first stator disposed at one surface of the rotor and including a first vibrating plate that is elastically deformable and a first...
8	Noise, Weight of stationary object, Weight of moving object, Power, Area of moving object, Area of stationary object	A vibration generator includes a tubular magnetic case, an end cover holding a first bearing metal, a second bearing metal held in a burring portion of the bottom plate of the case, a non-magnetic mov...
9	Stress or pressure, Force or torque, Loss of information, Power, Other harmful effects	An example method of reducing loads on a connection shaft includes disengaging a connection shaft from a motor-generator such that the connection shaft is not rotatably coupled to the motor-generator....



	generated by system, Energy used by stationary object	
10	Other harmful effects generated by system, Compatibility or connectability	Some embodiments of the present disclosure provide a rotor of an electric brushless motor configured to be light weight and prevent vibrations generated during an operation of the motor to be transfer...

表 4-9 為查詢第五道問題的檢索工程參數表結果，其中藍底色的為與檢索全文重疊的專利文本。第五句的查詢語句為 “An electric motor produces excessive vibration, causing fatigue failure.”，該語句沒有跟工程參數有匹配的字詞，其中內文中有與語句查詢匹配的字詞的可以查看圖 4-1 所示。

3. CONTENT RELEVANCE VERIFICATION:	
1. ID: 13222257	- Query terms found: ['motor'] (12.5% coverage)
2. ID: 13007199	- Query terms found: ['electric'] (12.5% coverage)
3. ID: 14124454	- Query terms found: ['electric'] (12.5% coverage)
4. ID: 12566292	- Query terms found: ['motor'] (12.5% coverage)
5. ID: 12303573	- Query terms found: ['electric'] (12.5% coverage)
6. ID: 12311515	- Query terms found: [] (0.0% coverage)
7. ID: 13763867	- Query terms found: [] (0.0% coverage)
8. ID: 14116041	- Query terms found: [] (0.0% coverage)
9. ID: 13047950	- Query terms found: ['motor'] (12.5% coverage)
10. ID: 12192059	- Query terms found: ['motor'] (12.5% coverage)

圖 4-1、查詢語句字詞匹配數

從結果觀察可知，即使查詢語句中未出現與工程參數直接相關的語詞，檢索器仍能解釋語句與參數之間的潛在語意關聯，並據此完成檢索任務。特別是在第六篇至第八篇推薦文本中，雖未明確提及如 electric 或 motor 等字詞，卻出現 vibration(黃底)與技術問題相關的語彙，顯示檢索器具備針對隱含技術問題進行語意推理與參數對應的能力，而非僅仰賴表層字詞的匹配。此能力使其能跨越特定領域，提供具多元性的推薦結果，展現出語義理解技術在工程領域檢索中的潛力。

## 第五章 結論與建議

隨著語言模型的更新速度從每年一次加快至每月甚至更頻繁，Ho et al. (2024) 研究指出其發展遠超過摩爾定律。面對資訊爆炸與技術快速演進，動態更新知識庫並有效查詢巨量文本資料成為關鍵課題。本研究以此為出發點，透過標註專利文本、建構專屬知識庫，進一步優化檢索效能並推薦符合創新需求之專利資料。

### 第一節 結論

本研究基於 TRIZ 歸納的工程參數，結合傳統與現代自然語言處理技術，針對專利文本進行 48 項工程參數的標註，以提升客觀性並降低人工判斷的誤差。同時，轉換為向量資料庫以優化檢索效率。採用文獻研究，運用組合式分類器計算相似度，再透過檢索器找出相關性最高的十篇專利文本。在不重新訓練語言模型的情況下，有效建構出一套融合 TRIZ 架構的專利資料庫系統。此外，本研究也透過不同的語句查詢進行資料分析與結果比較，驗證其可行性與實用性。

本研究主要有三項貢獻：

1. 本研究採用 Polikar. (2006) 所提出的分類器以及決策架構，建立一套專利文本工程參數的標註流程。
2. 本研究建構出在無需訓練與微調語言模型的前提下，提升語言模型訪問整體專利資料庫的效率。
3. 本研究架構了專利文本與工程參數關係的資料表，導入向量檢索技術的實作。

### 第二節 研究限制

#### 5.2.1 關鍵字詞庫的建立

本研究所建構之工程參數關鍵字詞庫由語言模型生成，雖已設定提示條件包含科學、工程技術與 TRIZ 等層面，尚須進一步驗證其標註準確性與涵蓋範圍，並視需求擴充詞彙，以提升標註精度與模型泛化能力。

### 5.2.2 分類器語文本權重

分類器語文本段落所設定之權重比例，尚未經由交叉實驗與參數調整進行系統性的驗證，故目前所採用之權重組合是否為最適配置，仍待後續量化分析與實證探討。

### 5.2.3 檢索器設計

本研究尚未針對檢索器進行效能優化與設計調整，導致查詢語句在回應速度上未明顯提升。考量向量檢索需與嵌入模型的維度相容性，若未來導入不同嵌入模型，需重新建置相對應維度之向量資料庫，以確其可行性與檢索準確性。

未來的研究方向可進一步探討如何將查詢語句透過 SBERT 或其他語言模型進行語意轉換，並將其對應至 TRIZ 所定義的 48 個工程參數維度。此方法不僅可提升查詢效率，降低檢索時間，同時也能使語句向參數的對應機制更加透明，強化整體系統的可視性與可控性。透過語意嵌入結構化對應於工程參數的維度，建立更具解釋力的語義檢索流程，提升模型在工程問題分析與推薦準確性上的表現。

### 5.2.4 標籤與檢索結果驗證

目前系統僅就詞彙頻率層級進行文本比對與檢索結果分析，尚未考慮語意層級的驗證方法。因此，系統是否有助於產品開發者提升創新表現，仍須仰賴其他方式（如設計問卷進行使用者評估）加以驗證其成效。

## 第三節 未來研究與建議

關於本研究雖然建立了一套能將專利文本向量化的資料，但仍有許多不足之處，待改進的地方如下：

1. 提升工程參數標註精準度與自動化機制：可進一步探索以多語言語言模型協助建立更具彈性且語意敏感的關鍵詞庫，提升模型在跨領域及跨語系專利文本中標註工程參數的準確性，並降低人工審閱負擔。
2. 分類器權重調整與最佳化實驗：未來可透過多種文本結構進行交叉實驗，分析不



同分類器參數配置對標註結果的影響，以出最適權重比例並提升整體模型效能。

3. 檢索器與向量維度調整策略：研究可進一步探討動態維度匹配演算法，使檢索器能根據不同嵌入模型自動調整向量資料庫結構，提升系統在替換嵌入模型時的適應性與查詢效率。另外，也可以設計客製化的檢索器，將查詢語句轉成 48 維度，即可直接進行相似度計算，提升檢索效能。
4. 語意層級驗證與使用者導向評估機制：未來可加入語意相似度分析工具，並結合創新指標或使用者問卷進行效益驗證，針對標註結果對創新開發者的實際幫助進行量化評估，提升其應用性與可控性。
5. TRIZ 工程參數與問題解構模型整合：可嘗試將 TRIZ 的矛盾矩陣、發明原則進一步納入檢索過程，讓使用者不僅能查詢相關專利，也能針對特定工程矛盾進行問題解決與創新推薦。



中 原 大 學

## 參考文獻

### 中文部分：

- 經濟部統計處 (2024)，我國發明專利申請以半導體技術領域為主，《產業經濟統計簡訊》，第 458 期
- 蔡元豪(2012)，萃智(TRIZ)演化趨勢專利分類系統建置—氣動打釘機之實證研究，國立臺北科技大學工業工程與管理學系碩士論文。
- 蔡明政(2021)，運用專利搜尋與文字探勘來探討產品設計之演化趨勢模式，未出版之碩士論文，中原大學工業與系統工程研究所，桃園縣。
- 謝凌翔(2021)，應用自然語言處理技術探討產品設計與專利推薦之研究，未出版之碩士論文，中原大學工業與系統工程研究所，桃園縣。
- 蔣昌錫(2023)，基於 TRIZ 之專利文本知識表達與特徵擷取研究，未出版之碩士論文，中原大學工業與系統工程研究所，桃園縣。

### 外文部分：

- Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613–620. <https://doi.org/10.1145/361219.361220>
- Goldberg, Y. (2015, October 6). A primer on neural network models for natural language processing [Manuscript draft]. arXiv. <http://arxiv.org/abs/1510.00726>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. arXiv preprint arXiv:1908.10084.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Arabzadeh, N., Yan, X., & Clarke, C. L. A. (2021). Predicting efficiency/effectiveness trade-offs for dense vs. sparse retrieval strategy selection. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21)*. <https://doi.org/10.48550/arXiv.2109.10739>
- Ri, R., Yamada, I., & Tsuruoka, Y. (2021). mLUKE: The power of entity representations in multilingual pretrained language models. arXiv preprint arXiv:2110.08151. <https://arxiv.org/abs/2110.08151>
- Glass, M., Rossiello, G., Chowdhury, M. F. M., Naik, A. R., Cai, P., & Gliozzo, A. (2022). Re2G: Retrieve, rerank, generate. arXiv preprint arXiv:2207.06300. <https://arxiv.org/abs/2207.06300>
- Bavarian, M., Jun, H., Tezak, N., Schulman, J., McLeavey, C., Tworek, J., & Chen, M.

- (2022). Efficient training of language models to fill in the middle. arXiv preprint arXiv:2207.14255. <https://arxiv.org/abs/2207.14255>
- Liang, P., Bommasani, R., Zhang, T., Holtzman, A., Wu, J., Kim, C., ... & Jurafsky, D. (2022). Holistic evaluation of language models. arXiv preprint arXiv:2211.06404. <https://arxiv.org/abs/2211.06404>
  - Ho, A., Besiroglu, T., Erdil, E., Owen, D., Rahman, R., Guo, Z. C., ... & Sevilla, J. (2024). Algorithmic progress in language models. arXiv preprint arXiv:2403.05812. <https://arxiv.org/abs/2403.05812>
  - Ghosh, S., Evuru, C. K. R., Kumar, S., Ramaneswaran, S., Aneja, D., Jin, Z., Duraiswami, R., & Manocha, D. (2024). A closer look at the limitations of instruction tuning. Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. arXiv:2402.05119v5. <https://arxiv.org/abs/2402.05119>
  - Chen, L., Song, Y., Ding, S., Sun, L., Childs, P., & Zuo, H. (2024). TRIZ-GPT: An LLM-augmented method for problem-solving. Proceedings of the ASME 2024 International Design Engineering Technical Conference and Computers and Information in Engineering Conference (IDETC-CIE2024), Washington, DC, USA. arXiv:2408.05897. <https://arxiv.org/abs/2408.05897>
  - Jeong, S., Baek, J., Cho, S., Hwang, S. J., & Park, J. C. (2024). Adaptive-RAG: Learning to adapt retrieval-augmented large language models through question complexity. arXiv preprint arXiv:2403.14403. <https://arxiv.org/abs/2403.14403>
  - Fan, W., Wang, S., Ding, Y., Li, H., Ning, L., Yin, D., Chua, T.-S., & Li, Q. (2024). A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. arXiv preprint arXiv:2405.06211. <https://arxiv.org/abs/2405.06211>
  - Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997. <https://arxiv.org/abs/2312.10997>
  - Rackauckas, Z. (2024). RAG-Fusion: A new take on retrieval-augmented generation. arXiv preprint arXiv:2402.03367. <https://arxiv.org/abs/2402.03367>
  - Zorik, A., Liu, J., & Khashabi, D. (2024). Does fine-tuning LLMs on new knowledge encourage hallucinations? arXiv preprint arXiv:2402.06691. <https://arxiv.org/abs/2402.06691>
  - Mann, D. (2002a). Evolving the inventive principles. The TRIZ journal.
  - Mann, D. (2002b). Hands on systematic innovation. The TRIZ journal.
  - Miller, D. (2019). Leveraging BERT for extractive text summarization on lectures. arXiv preprint arXiv:1906.04165.
  - Ehghaghi, M. (2024, June 25). The hidden obstacles of domain adaptation in LLMs. Arcee AI Blog. <https://www.arcee.ai/blog/the-hidden-obstacles-of-domain-adaptation-in-llms>
  - Koshaev, A., & Petrova, I. (2015). TRIZ-based analysis of patent contradictions. In M. B. Horowitz (Ed.), TRIZ – Systematic innovation in manufacturing (pp. 379–390). Springer. [https://doi.org/10.1007/978-3-319-13987-6\\_32](https://doi.org/10.1007/978-3-319-13987-6_32)

- Lin, C.-Y. (2006). A study of the contradiction analysis and resolution strategies based on TRIZ theory [Master's thesis, National Chiao Tung University]. National Chiao Tung University Institutional Repository. <https://ir.lib.nycu.edu.tw/handle/11536/77077>
- LangChain. (n.d.). LangChain: The open framework for building LLM applications. <https://www.langchain.com/>
- Hugging Face. (n.d.). Hugging Face: The AI community building the future. <https://huggingface.co/>
- United States Patent and Trademark Office. (n.d.). Patent Public Search. Retrieved from <https://ppubs.uspto.gov/pubwebapp/>
- OpenAI. (n.d.). OpenAI API documentation: Introduction. Retrieved from <https://platform.openai.com/docs/api-reference/introduction>



## ■ 附錄一 48 工程參數英文說明

No.	Engineering Parameter	Definition
1	Weight of moving object	The mass of the object, in a gravitational field. The force that the body exerts on its support or suspension.
2	Weight of stationary object	The mass of the object, in a gravitational field. The force that the body exerts on its support or suspension, or on the surface on which it rests.
3	Length/Angle of moving object	Any one linear dimension, not necessarily the longest, is considered a length.
4	Length/Angle of stationary object	Any one linear dimension, not necessarily the longest, is considered a length.
5	Area of moving object	A geometrical characteristic described by the part of a plane enclosed by a line. The part of a surface occupied by the object. OR the square measure of the surface, either internal or external, of an object.
6	Area of stationary object	A geometrical characteristic described by the part of a plane enclosed by a line. The part of a surface occupied by the object. OR the square measure of the surface, either internal or external, of an object.
7	Volume of moving object	The cubic measure of space occupied by the object. Length x width x height for a rectangular object, height x area for a cylinder, etc.
8	Volume of stationary object	The cubic measure of space occupied by the object. Length x width x height for a rectangular object, height x area for a cylinder, etc.
9	Shape	The external contours, appearance of a system.
10	Amount of substance	The number or amount of a system' s materials, substances, parts or subsystems which might be changed fully or partially, permanently or temporarily.
11	Amount of information	The amount, quantity or number of a system' s informational resources. Information should be interpreted in its most generic form to include any form of information that might be passed between two or more objects or systems.
12	Duration of action of moving object	The time that the object can perform the action. Service life. Mean time between failure is a measure of the duration of action. Also, durability.

13	Duration of action of stationary object	The time that the object can perform the action. Service life. Mean time between failure is a measure of the duration of action. Also, durability.
14	Speed	The velocity of an object; the rate of a process or action in time.
15	Force/torque	Force measures the interaction between systems. In Newtonian physics, force = mass x acceleration. In TRIZ, force is any interaction that is intended to change an object's condition.
16	Energy used by moving object	The measure of the object's capacity for doing work. In classical mechanics, Energy is the product of force x distance. This includes the use of energy provided by the super - system (such as electrical energy or heat.) Energy required to do a particular job.
17	Energy used by stationary object	The measure of the object's capacity for doing work. In classical mechanics, Energy is the product of force x distance. This includes the use of energy provided by the super - system (such as electrical energy or heat.) Energy required to do a particular job.
18	Power	The time rate at which work is performed. The rate of use of energy.
19	Stress/pressure	Force per unit area. Also, tension.
20	Strength	The extent to which the object is able to resist changing in response to force. Resistance to breaking.
21	Stability	The wholeness or integrity of the system; the relationship of the system's constituent elements. Wear, chemical decomposition, and disassembly are all decreases in stability. Increasing entropy is decreasing stability.
22	Temperature	The thermal condition of the object or system. Loosely includes other thermal parameters, such as heat capacity, that affect the rate of change of temperature.
23	Illumination intensity	Light flux per unit area, also any other illumination characteristics of the system such as brightness, light quality, etc.
24	Function efficiency	The efficiency with which a system performs a specific task, characterized by the amount of time, energy, or other resources required. It is an important measure of a system's performance, as it determines the cost and effectiveness of the system.
25	Loss of substance	Partial or complete, permanent or temporary, loss of some of a system's materials, substances, parts or subsystems.

26	Loss of time	Time is the duration of an activity. Improving the loss of time means reducing the time taken for the activity. 'Cycle time reduction' is a common term.
27	Loss of energy	Use of energy that does not contribute to the job being done. See 19. Reducing the loss of energy sometimes requires different techniques from improving the use of energy, which is why this is a separate category.
28	Loss of information	Partial or complete, permanent or temporary, loss of data or access to data in or by a system. Frequently includes sensory data such as aroma, texture, etc.
29	Noise	Any unwanted sound produced by a system, which can have negative effects on human hearing or other systems. Noise can be caused by various factors, such as mechanical vibrations or electrical interference, and can be measured in decibels (dB).
30	Harmful emissions	Pollutants or other substances released by a system that can have negative impacts on the environment or human health. These emissions can come in the form of gases, particulates, or other materials, and can be regulated by government agencies to protect public health and the environment.
31	Other harmful effects generated by system	Any negative impact produced by a system, including effects on other systems. These effects can include physical damage, interference with other systems, or negative impacts on the environment.
32	Adaptability/versatility	The extent to which a system/object positively responds to external changes. Also, a system that can be used in multiple ways for under a variety of circumstances.
33	Compatibility/connectability	The ability of a system to work seamlessly with other systems or devices. This can include the ability to exchange data or interact with other systems, as well as the ability to physically connect to other devices or systems.
34	Ease of operation	Simplicity: The process is not easy if it requires a large number of people, large number of steps in the operation, needs special tools, etc. 'Hard' processes have low yield and 'easy' process have high yield; they are easy to do right.
35	Reliability/robustness	A system's ability to perform its intended functions in predictable ways and conditions.
36	Repairability	Quality characteristics such as convenience, comfort, simplicity, and time to repair faults, failures or defects in a system.

37	Security	Measures in place to protect a system from unauthorized access or operation. This can include physical security measures, such as locks or barriers, as well as cybersecurity measures, such as encryption or authentication.
38	Safety/ Vulnerability	The safety of a system and its susceptibility to accidents or attacks. This can include measures to prevent accidents or injuries, as well as measures to protect against intentional attacks or breaches.
39	Aesthetics/ appearance	The visual appeal and design of a system. This can include factors such as color, shape, and overall design, as well as how well the system integrates with its surroundings.
40	Other harmful effects acting on the system	Any negative impact on a system from external factors, such as environmental conditions. These can include physical impacts, such as extreme temperatures or vibrations, as well as chemical or biological factors.
41	Manufacturability	The degree of facility, comfort or effortlessness in manufacturing or fabricating the object/system.
42	Manufacture precision/ consistency	The extent to which the actual characteristics of the system or object match the specified or required characteristics.
43	Automation	The extent to which a system or object performs its functions without human interface. The lowest level of automation is the use of a manually operated tool. For intermediate levels, humans program the tool, observe its operation, and interrupt or re-program as needed. For the highest level, the machine senses the operation needed, programs itself and monitors its own operations.
44	Productivity	The number of functions or operations performed by a system per unit time. The time for a unit function or operation. The output per unit time, or the cost per unit output.
45	System complexity	The number of components and interactions within a system. A complex system may have many parts that must work together in order to function properly, which can increase the difficulty of designing, operating, and maintaining the system.
46	Control complexity	The level of difficulty in operating and controlling a system, including the number of steps and decisions required. A complex control system may have many controls or settings that must be adjusted in order to properly operate the system, which can increase the difficulty of using and maintaining the



		system. This can also affect the system' s reliability and performance.
47	Ability to detect or measure	The capability to identify and quantify aspects of a system, such as performance or output. This can include the use of sensors or other measurement tools to monitor and assess the system' s performance.
48	Measurement precision	The closeness of the measured value to the actual value of a property of a system. Reducing the error in a measurement increases the accuracy of the measurement.



中 原 大 學

## 附錄二 48 工程參數關鍵字詞表

工程參數代號	關鍵字詞
1	["weight", "mass", "heavy", "light", "moving", "mobile", "dynamic"]
2	["weight", "mass", "heavy", "light", "stationary", "fixed", "static"]
3	["length", "angle", "long", "short", "dimension", "angular", "moving", "mobile"]
4	["length", "angle", "long", "short", "dimension", "angular", "stationary", "fixed"]
5	["area", "surface", "coverage", "moving", "mobile"]
6	["area", "surface", "coverage", "stationary", "fixed"]
7	["volume", "capacity", "size", "moving", "mobile"]
8	["volume", "capacity", "size", "stationary", "fixed"]
9	["shape", "form", "geometry", "configuration", "design", "contour"]
10	["amount", "substance", "quantity", "material", "matter", "mass"]
11	["amount", "information", "data", "knowledge", "content", "details"]
12	["duration", "time", "period", "lifetime", "moving", "action"]
13	["duration", "time", "period", "lifetime", "stationary", "action"]
14	["speed", "velocity", "fast", "slow", "rate", "acceleration"]
15	["force", "torque", "push", "pull", "tension", "compression", "twist"]
16	["energy", "power", "consumption", "efficiency", "moving", "dynamic"]
17	["energy", "power", "consumption", "efficiency", "stationary", "static"]
18	["power", "energy", "electrical", "mechanical", "output", "wattage"]
19	["stress", "pressure", "strain", "load", "mechanical", "compression"]
20	["strength", "durability", "robust", "strong", "weak", "mechanical"]
21	["stability", "stable", "unstable", "equilibrium", "balance", "steady"]
22	["temperature", "heat", "cold", "thermal", "heating", "cooling"]
23	["illumination", "light", "brightness", "lighting", "visibility"]
24	["function", "efficiency", "performance", "effectiveness", "capability"]
25	["loss", "waste", "substance", "material", "leakage"]
26	["loss", "time", "delay", "efficiency", "speed"]
27	["loss", "waste", "energy", "efficiency", "dissipation"]
28	["loss", "information", "data", "signal", "communication"]
29	["noise", "sound", "vibration", "acoustic", "disturbance"]

30	["harmful", "emissions", "pollution", "toxic", "waste", "byproduct"]
31	["harmful", "effects", "generated", "system", "internal", "byproduct"]
32	["adaptability", "versatility", "flexible", "adjustable", "modular"]
33	["compatibility", "connectability", "interoperability", "standard", "interface"]
34	["ease", "operation", "use", "user", "interface", "simple"]
35	["reliability", "robustness", "dependable", "consistent", "failure"]
36	["repairability", "repair", "maintenance", "service", "fix"]
37	["security", "secure", "protection", "encryption", "access"]
38	["safety", "vulnerability", "hazard", "risk", "danger"]
39	["aesthetics", "appearance", "beauty", "visual", "design"]
40	["harmful", "effects", "acting", "external", "damage"]
41	["manufacturability", "manufacture", "production", "fabrication", "easy"]
42	["manufacture", "precision", "consistency", "tolerance", "production", "quality"]
43	["automation", "automatic", "manual", "control", "extent"]
44	["productivity", "output", "efficiency", "throughput", "performance"]
45	["system", "complexity", "simple", "complicated", "device"]
46	["control", "complexity", "simple", "complicated", "management"]
47	["ability", "detect", "measure", "sensing", "monitoring", "detection"]
48	["measurement", "precision", "accuracy", "error", "calibration"]