中 原 大 學

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

陳崇恩

中

華

民

國

114

年

7

月

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

**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](#_Toc127527354)

[Abstract II](#_Toc127527355)

[致 謝 III](#_Toc127527356)

[目 錄 IV](#_Toc127527357)

[圖 目 錄 VI](#_Toc127527358)

[表 目 錄 VIII](#_Toc127527359)

[第一章 緒 論 1](#_Toc127527360)

[第一節 研究動機 1](#_Toc127527361)

[第二節 研究目的 3](#_Toc127527362)

[第三節 研究流程 4](#_Toc127527363)

[第二章 文獻探討 6](#_Toc127527364)

[第一節 萃智創新發明問題解決理論 6](#_Toc127527365)

[第二節 功能屬性分析 7](#_Toc127527366)

[2.2.1 功能 8](#_Toc127527367)

[2.2.2 屬性 9](#_Toc127527368)

[2.2.3 四十發明原則 10](#_Toc127527369)

[第三節 知識表示 12](#_Toc127527370)

[2.3.1 智慧財產權 12](#_Toc127527371)

[2.3.2 知識庫建構 14](#_Toc127527372)

[2.3.3 知識應用 18](#_Toc127527373)

[第四節 專利探勘 20](#_Toc127527374)

[第五節 自然語言處理 23](#_Toc127527375)

[2.5.1 Transformer套件模型 24](#_Toc127527376)

[2.5.2 PyTorch Lighting框架 25](#_Toc127527377)

[2.5.3 BERT模型 25](#_Toc127527378)

[2.5.4 GPT - 3 27](#_Toc127527379)

[第六節 語義相似度 28](#_Toc127527380)

[2.6.1 TF - IDF 29](#_Toc127527381)

[2.6.2 Word2Vec模型 30](#_Toc127527382)

[2.6.3 Doc2Vec模型 31](#_Toc127527383)

[2.6.4 餘弦相似性 33](#_Toc127527384)

[第三章 研究方法 35](#_Toc127527385)

[第一節 建構功能屬性衍伸字詞庫 36](#_Toc127527386)

[第二節 轉化為向量並判斷相似度 38](#_Toc127527387)

[3.2.1判斷功能屬性關鍵字詞庫與發明原則建立相似度 40](#_Toc127527388)

[第三節 發明原則與功能屬性相似度分類 42](#_Toc127527389)

[3.3.1屬性功能關鍵字應用 44](#_Toc127527390)

[第四節 專利文本集選取訓練內容 44](#_Toc127527391)

[3.4.1判斷屬性關鍵字詞庫與專利文本建立相似度 45](#_Toc127527392)

[3.4.2發明原則&屬性功能少量標籤提取 46](#_Toc127527393)

[第五節 基於BERT模型的計算方式 47](#_Toc127527394)

[3.5.1自注意力機制 49](#_Toc127527395)

[3.5.2餘弦相似度 50](#_Toc127527396)

[第四章 研究結果 52](#_Toc127527397)

[第一節 Doc2Vec協助建立相關性分析結果 52](#_Toc127527398)

[4.1.1驗證發明原則與功能屬性資料集彙整 52](#_Toc127527399)

[4.1.2驗證發明原則與專利資料集彙整 55](#_Toc127527400)

[4.1.3驗證功能屬性與專利資料集彙整 57](#_Toc127527401)

[第二節 BERT專利文本與發明原則分析結果 57](#_Toc127527402)

[4.2.1分類效果驗證 58](#_Toc127527403)

[第三節 BERT專利文本與功能屬性參數辨識結果 59](#_Toc127527404)

[4.3.1功能參數分類效果驗證 60](#_Toc127527405)

[4.3.2屬性參數分類效果驗證 61](#_Toc127527406)

[第四節 專利標籤與產品設計之模式分析 62](#_Toc127527407)

[第五章 結論與建議 67](#_Toc127527408)

[第一節 結論 67](#_Toc127527409)

[第二節 研究限制 68](#_Toc127527410)

[第三節 未來研究與建議 68](#_Toc127527411)

[參考文獻 70](#_Toc127527412)

[附錄一 40發明原則說明 74](#_Toc127527413)

[附錄二 40發明原則功能屬性分項列表 79](#_Toc127527414)

[附錄三 功能字詞庫 84](#_Toc127527415)

[附錄四 屬性字詞庫 86](#_Toc127527416)

**圖 目 錄**

[圖1 - 1、研究流程圖 4](#_Toc127528014)

[圖2 - 1、TRIZ解題流程 7](https://d.docs.live.net/ae73d235492c85db/110計畫研究(908-2)/學位考試/基於TRIZ之專利文本知識表達與特徵擷取研究(0216).docx#_Toc127527997)

[圖2 - 2、專利功能六要素 (葉子諒, 2017) 8](#_Toc127527998)

[圖2 - 3、PSG - TCO模型描述 17](#_Toc127527999)

[圖2 - 4、自動知識挖掘方法的概述 18](#_Toc127528000)

[圖2 - 5、PSG - TCO實例圖 18](https://d.docs.live.net/ae73d235492c85db/110計畫研究(908-2)/學位考試/基於TRIZ之專利文本知識表達與特徵擷取研究(0216).docx#_Toc127528001)

[圖2 - 6、專利文本的表示模型 19](https://d.docs.live.net/ae73d235492c85db/110計畫研究(908-2)/學位考試/基於TRIZ之專利文本知識表達與特徵擷取研究(0216).docx#_Toc127528002)

[圖2 - 7、文字探勘流程 21](#_Toc127528003)

[圖2 - 8、研究總體流程 22](#_Toc127528004)

[圖2 - 9、Transformer架構(Vaswani et al., 2017) 24](#_Toc127528005)

[圖2 - 10、BERT在預訓練時需要完成的兩個任務 27](#_Toc127528006)

[圖2 - 11、詞袋分析步驟 29](#_Toc127528007)

[圖2 - 12、Word2Vec的兩種模型：CBOW與Skip - gram 31](#_Toc127528008)

[圖2 - 13、PV - DM模型(Le & Mikolov, 2014) 32](#_Toc127528009)

[圖2 - 14、PV - DBOW模型(Le & Mikolov, 2014) 33](#_Toc127528010)

[圖2 - 15、夾角θ與餘弦值關係 33](#_Toc127528011)

[圖2 - 16、餘弦相似度計算方法 34](#_Toc127528012)

[圖3 - 1、研究方法流程圖 36](#_Toc127527975)

[圖3 - 2、PV - DM模型結構 38](#_Toc127527976)

[圖3 - 3、Doc2Vec維度說明 39](#_Toc127527977)

[圖3 - 4、PV - DBOW模型結構 39](#_Toc127527978)

[圖3 - 5、Doc2vec模型訓練流程圖 40](#_Toc127527979)

[圖3 - 6、40發明原則內文 41](#_Toc127527980)

[圖3 - 7、輸出與功能 - 改變相位有關的前10個發明原則虛擬碼 41](#_Toc127527981)

[圖3 - 8、輸出與屬性 - 物體穩定性有關的前10個發明原則虛擬碼 42](#_Toc127527982)

[圖3 - 9、句子中單字長度統計圖 45](#_Toc127527983)

[圖3 - 10、Doc2vec模型訓練流程圖 46](#_Toc127527984)

[圖3 - 11、輸出與屬性 - 顏色前10個有關專利虛擬碼 46](#_Toc127527985)

[圖3 - 12、專利號少量標籤順序 47](#_Toc127527986)

[圖3 - 13、Transformer架構 48](#_Toc127527987)

[圖3 - 14、自注意力機制運算方式 50](#_Toc127527988)

[圖3 - 15、多標頭自注意力機制運算方式 50](#_Toc127527989)

[圖4 - 1、Doc2Vec虛擬碼說明 53](#_Toc127527963)

[圖4 - 2、Doc2Vec虛擬碼說明 53](#_Toc127527964)

[圖4 - 3、Doc2Vec虛擬碼說明 54](#_Toc127527965)

[圖4 - 4、專利本地數據庫 55](#_Toc127527966)

[圖4 - 5、發明原則與專利相似度比對 56](#_Toc127527967)

[圖4 - 6、發明原則專利標籤 58](#_Toc127527968)

[圖4 - 7、功能屬性專利標籤 60](#_Toc127527969)

**表 目 錄**

[表2 - 1、屬性改變方式 9](#_Toc127449832)

[表2 - 2、屬性兩種定義 10](#_Toc127449833)

[表2 - 3、40發明原則(部分節錄自林瑞庭, 2011) 10](#_Toc127449834)

[表2 - 4、智慧財產權兩大體系 13](#_Toc127449835)

[表2 - 5、專利文獻弊端(林瑞庭, 2011) 14](#_Toc127449836)

[表2 - 6、IPC代碼結構示例 20](#_Toc127449837)

[表2 - 7、其研究中所使用的術語 22](#_Toc127449838)

[表2 - 8、典型的政府利益聲明 23](#_Toc127449839)

[表3 - 1、功能字詞庫 37](#_Toc127449823)

[表3 - 2、屬性字詞庫 38](#_Toc127449824)

[表3 - 3、40發明原則功能屬性分項列表 43](#_Toc127449825)

[表4 - 1、40發明原則功能屬性分項列表 55](#_Toc127449813)

[表4 - 2、專利少量標籤 56](#_Toc127449814)

[表4 - 3、功能屬性少量標籤 57](#_Toc127449815)

[表4 - 4、發明原則分類結果 58](#_Toc127449816)

[表4 - 5、功能參數分類結果 60](#_Toc127449817)

[表4 - 6、屬性參數分類結果 61](#_Toc127449818)

[表4 - 7、驗證發明原則標籤 62](#_Toc127449819)

[表4 - 8、驗證功能參數標籤 63](#_Toc127449820)

[表4 - 9、驗證屬性參數標籤 63](#_Toc127449821)

1. **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，但往往散布於不同專利中。

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。

1. **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Genrish 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種工程參數

|  |  |
| --- | --- |
| 1. Weight of moving object | 1. Weight of stationary object |
| 1. Length or angle of moving object | 1. Length or angle of stationary object |
| 1. Area of moving object | 1. Area of stationary object |
| 1. Volume of moving object | 1. Volume of stationary object |
| 1. Shape | 1. Amount of substance |
| 1. Amount of information | 1. Duration of action of moving object |
| 1. Duration of action of stationary object | 1. Speed |
| 1. Force or torque | 1. Energy used by moving object |
| 1. Energy used by stationary object | 1. Power |
| 1. Stress or pressure | 1. Strength |
| 1. Stability | 1. Temperature |
| 1. Illumination intensity | 1. Function efficiency |
| 1. Loss of substance | 1. Loss of time |
| 1. Loss of energy | 1. Loss of information |
| 1. Noise | 1. Harmful emissions |
| 1. Other harmful effects generated by system | 1. Adaptability or versatility |
| 1. Compatibility or connectability | 1. Ease of operation |
| 1. Reliability or robustness | 1. Repairability |
| 1. Security | 1. Safety or vulnerability |
| 1. Aesthetics or appearance | 1. Other Harmful Effects Acting On the System |
| 1. Manufacturability | 1. Manufacture precision or consistency |
| 1. Automation | 1. Productivity |
| 1. System complexity | 1. Control complexity |
| 1. Ability to detect or measure | 1. 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汙染物，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。

|  |  |
| --- | --- |
| 一張含有 文字, 圖表, 字型, 螢幕擷取畫面 的圖片  AI 產生的內容可能不正確。  圖2-2、siamese 架構圖(Reimers et al. 2019) | 一張含有 文字, 圖表, 行, 螢幕擷取畫面 的圖片  AI 產生的內容可能不正確。  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 與 ，以及其差的絕對值 進行串接並與可訓練的權重 相乘，如公式(1)：

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

1. 回歸目標函數：計算兩個句子嵌入向量 與 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。
2. triplet目標函數：給定一組錨點句子 、正樣本 及負樣本 ，透過三元組損失函數(triplet loss) 使 與 之間的距離小於 與 的距離，並最小化以下的損失函數，如公式(2)：

其中 分別為 句子的嵌入向量， 為距離度量， 為邊界值，確保 比 更接近 至少 以上。如此，就可以微調模型並產生具有語意的句子遷入向量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 、 代表不同的文本：

**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。

一張含有 文字, 圖表, 方案, 工程製圖 的圖片

AI 產生的內容可能不正確。

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。

A diagram of a software

AI-generated content may be incorrect.

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 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。

A diagram of a diagram

AI-generated content may be incorrect.

圖2-6、RAG詳細架構 (Lewis et al. 2020)

**2.5.1 檢索器**

由RAG以上的架構，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。

**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。

A screenshot of a computer

AI-generated content may be incorrect.

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。

1. **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

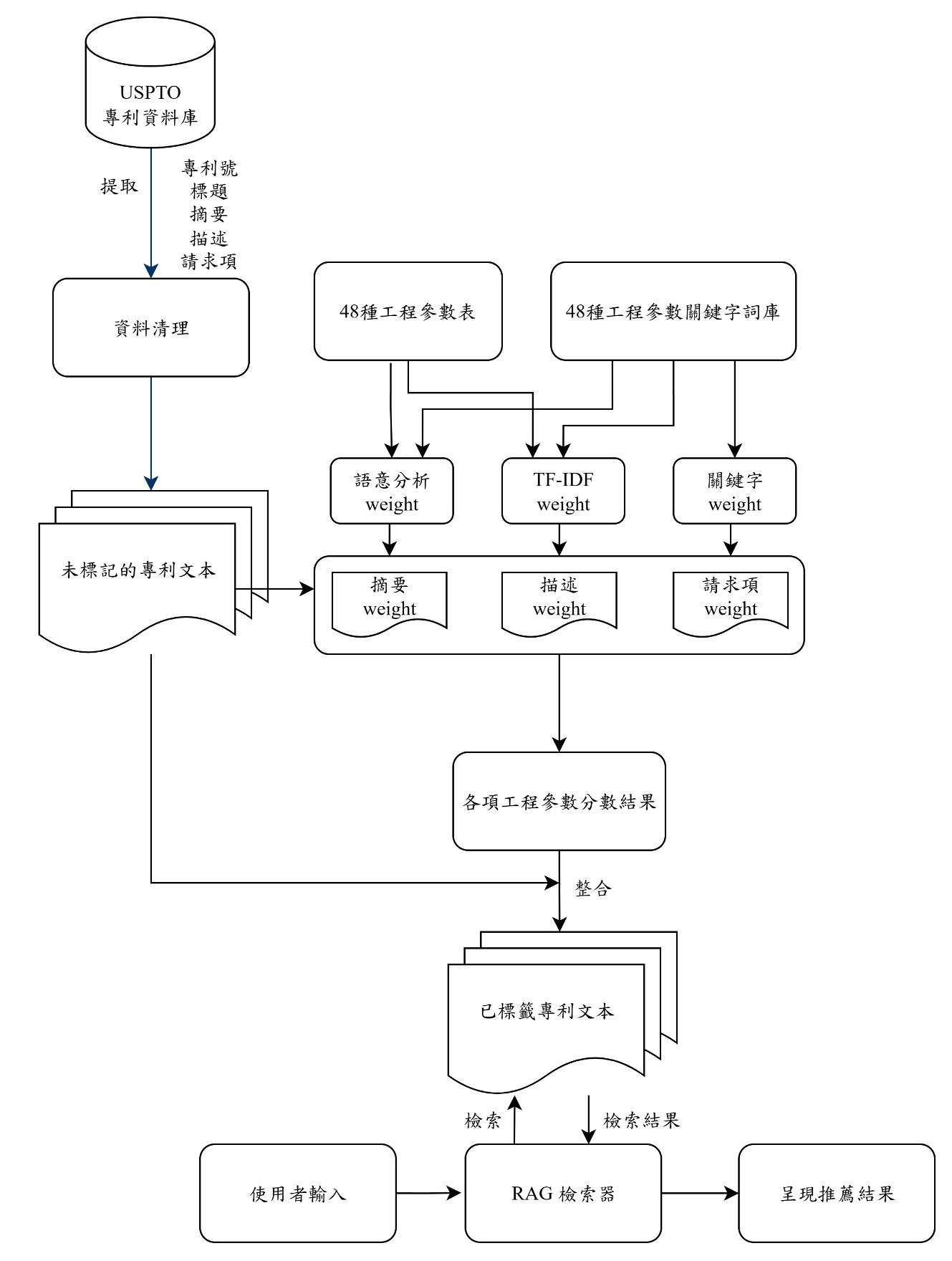


圖3 - 1、研究方法流程圖

**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。

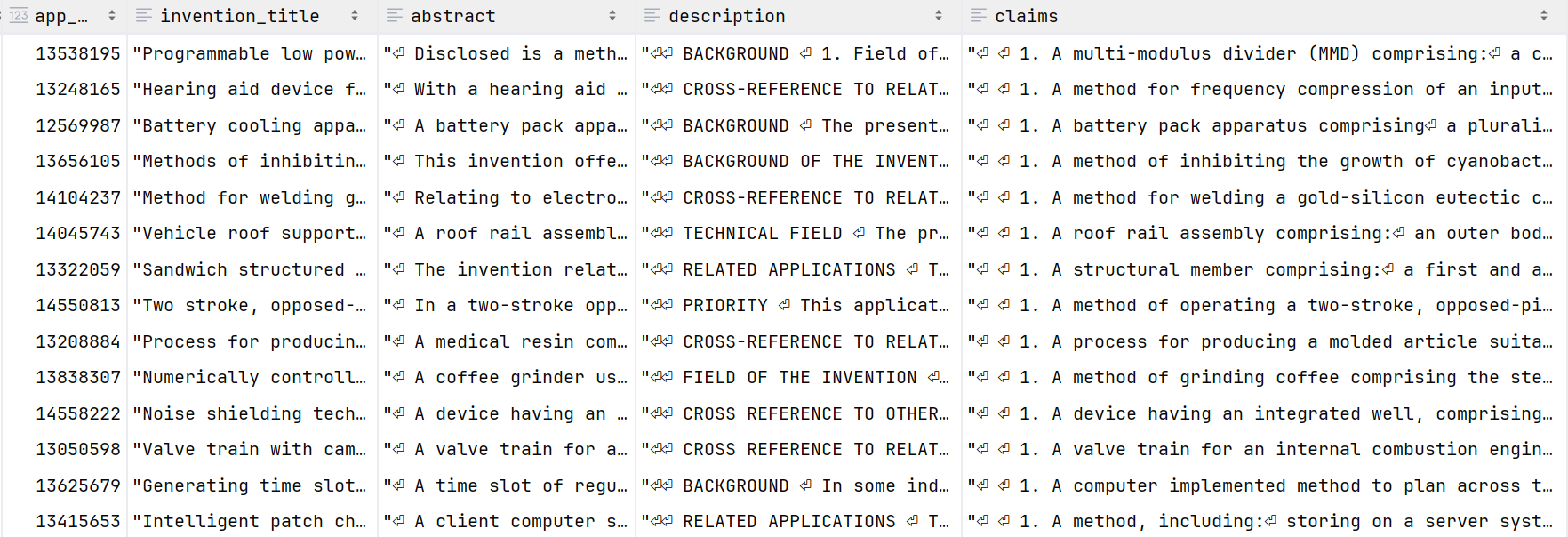


圖3-2、提取欄位後的資料

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。

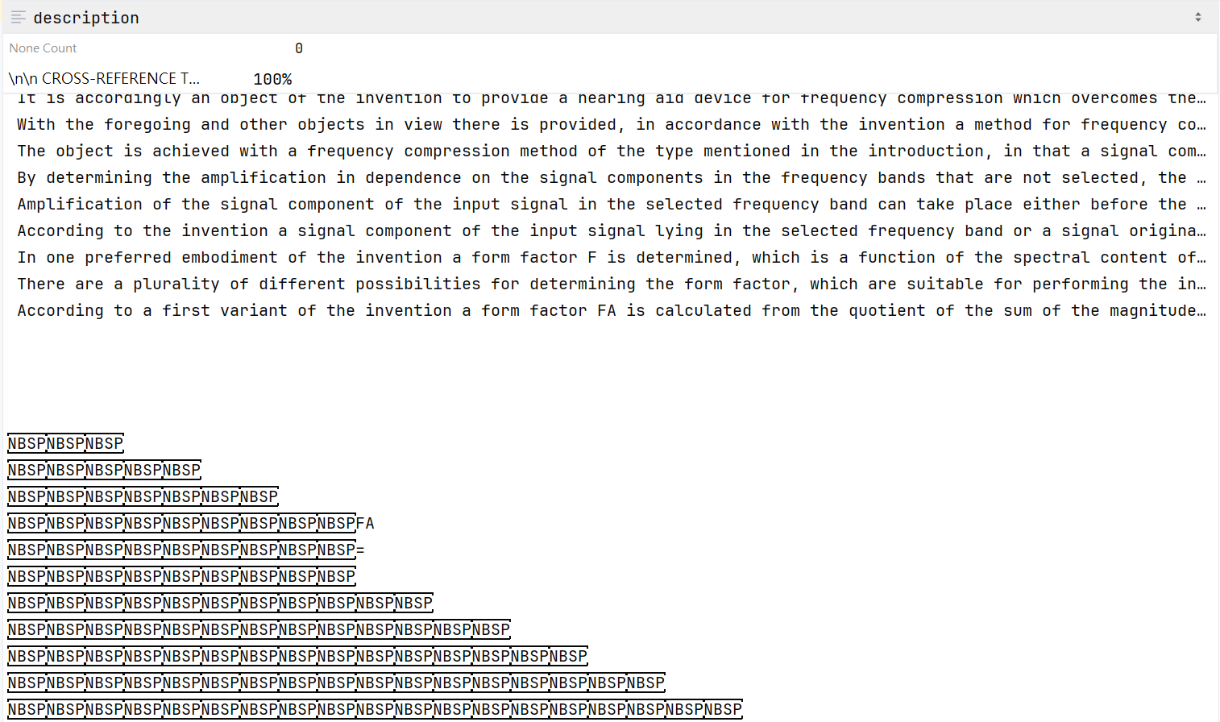


圖3-3、原始文本內容

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)

一張含有 文字, 字型, 紙張, 文件 的圖片

AI 產生的內容可能不正確。

圖3-4、進行處理後的文本內容

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

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。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

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：

* 為輸入的原始文本
* 為關鍵字
* 函數 代表關鍵字在文本 中出現的次數

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：

* ，為分類器，為工程參數 的分數
* 為權重

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。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

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 RAG檢索器**

經由USPTO的專利資料集，當中選取中重要項目，例如：ID、摘要、專利描述以及專利要求物，這些項目集成之後會形成專利本地數據庫，一共是5,489筆資料。然而，本研究需要將文本數據轉換為數字表示，Transformer模型一次不能處理超過512個單詞，能夠符合上述的欄位項目，只有摘要的部分能夠使用，專利描述的部分單詞幾乎大於800，所以無法使用，也因此本研究將限制在前400個字。然而，為了確定專利摘要是否皆小於400字，圖3 - 9的X軸為句子單字數統計，Y軸為出現次數，可以看出150字的句子出現次數最多，所以把字數限制在前400字是合理的。

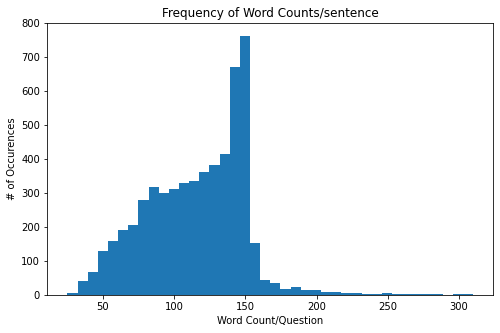


圖3 - 9、句子中單字長度統計圖

**3.4.1****判斷屬性關鍵字詞庫與專利文本建立相似度**

如何找出專利文本摘要與功能屬性相似性，運用Doc2ve客觀比較專利摘要與功能屬性相似之處，模型訓練流程如圖3 - 10。首先，此次建立相關性的實驗數據庫是專利摘要的內文全文CSV檔，一共是5,489筆專利資料，讓Doc2vec讀取專利ID與摘要文本進行訓練，因為內文資料量較大需要較久時間進行模型訓練，再使用屬性的關鍵字詞庫去做相似度比對，每個屬性字詞列出前10大相似度與其相近的專利號碼與摘要文本如圖3 - 11。

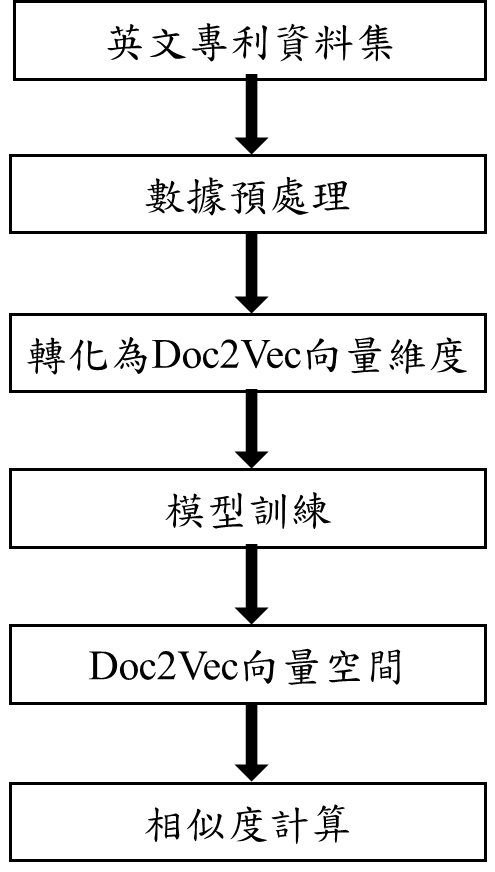


圖3 - 10、Doc2vec模型訓練流程圖

*Create an empty array called "questions".*

*Loop through the range from 0 to the number of tokens in the document.*

*Append the vector for each token to the "questions" array.*

*Convert the "questions" array into a numpy array.*

*Create a variable "text" and assign it the value "****color****".*

*Tokenize the text and remove stopwords, and store the result in "filtered\_tokens".*

*Print "filtered\_tokens".*

*Calculate the cosine similarity between the vector of "filtered\_tokens" and the "questions" array and store the result in "similarity".*

*Store the indices of the top 10 values in "similarity" in "top\_n".*

*Print the indices in "top\_n".*

*Loop through the range from 0 to the number of elements in "top\_n".*

*Print the value of "similarity[0][i]" and "corpus[i]".*

圖3 - 11、輸出與屬性 - 顏色前10個有關專利虛擬碼

**3.4.****2****發明原則&屬性功能少量標籤提取**

至於如何進行少量標籤提取，當然必需載入資料庫才能發現其關係，運用上述段落所述的方式進行相似度比對，優先比對屬性字詞庫，選取其相似度數字高的專利，給予其標籤，總共有213筆專利標籤，其對應193個屬性字詞，再從213筆專利當中比對相似度高的功能字詞，功能字詞共有56個，最後對應其發明原則，發明原則共有40個如圖3 - 12。

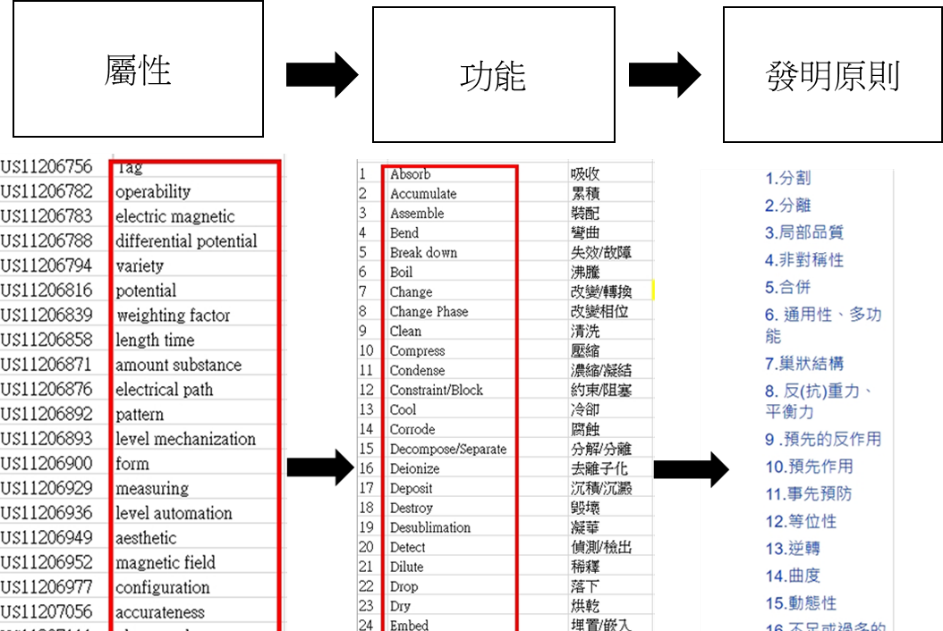


圖3 - 12、專利號少量標籤順序

**第五節 整合RAG流程**

BERT是傳統語言模型的一種變形，其全名為(Bidirectional Encoder Representations from Transformers)，而語言模型所做的事就是在給定一些詞彙的前提下，去估算下一個詞彙出現的機率分布。在給定t個在字典裡的詞彙，語言模型需要去估算第t+1個詞彙的機率分布P，其中可以是任何字在字典，公式(1)如下:

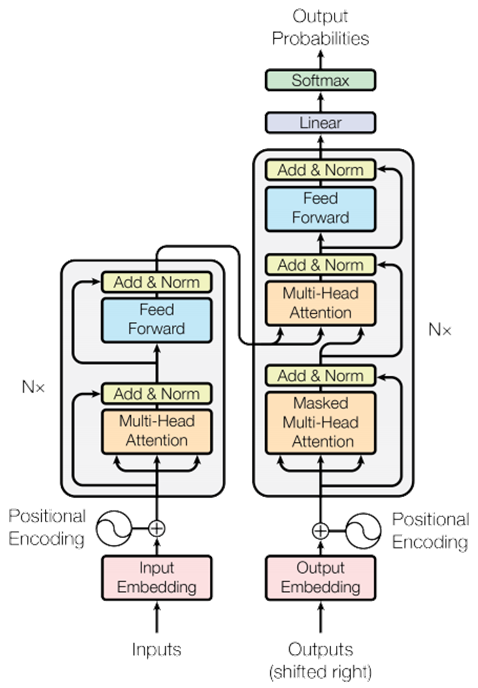


圖3 - 13、Transformer架構

Vaswani et al. (2017)Transformer遵循這一總體結構，使用堆疊的自注意和點對點的全連接層，用於編碼器，圖3 - 13的左半部和右半部分別顯示了編碼器和解碼器的堆疊自注意和逐點連接層。

編碼器是由N=6個相同的層組成的堆疊，每個layer有兩個子層，第一層是一個多標頭的自注意機制，第二層是一個簡單且按位置排列的全連接前饋網路，接下來每兩個子層周圍使用一個殘差連接，然後進行層的標準化，也就是說每個子層的輸出是LayerNorm(x+sublayer(x))，其中sublayer(x)是子層本身實現的函數，為了方便這些殘差連接，模型中的所有子層以及嵌入層，都會產生維數為=512的輸出。

解碼器也是由N=6個相同層組成的堆疊，除每個編碼器層中的兩個子層外，解碼器還插入第三個子層，該子層對編碼器堆疊的輸出執行多標頭注意，概念與編碼器類似，我們在每個子層周圍使用殘差連接，然後進行層的標準化，還修改解碼器堆疊中的自注意力子層，以防止位置注意到後續位置。

由上述可以得知Transformer遵循編碼器和解碼器這兩個構造，然而BERT模型是屬於左半部的編碼器構造，透過BERT模型中的自注意力機制找出功能屬性以及發明原則與專利之間的相似度關係，至於要如何讓BERT清楚的明瞭一篇專利的摘要要給予它什麼知識結構，這個就必須於3.4.2節中的少量自定義標籤中，是如何給予的知識結構雛形，所以總結來說一篇專利必須擁有三個標籤：1.發明原則標籤2.功能標籤3.屬性標籤。

**3.5.1****自注意力機制**

Transformer克服了執行效能過慢的問題，提出自注意力機制(Self-Attention mechanism)，能夠平行計算出所有輸出，況且上述有提到自注意力機制，那麼其明確的計算方式如圖3 - 13。

首先輸入向量被表徵為Q、K、V向量(Vector)，K是鍵向量(Key Vector)，為編碼器隱藏層狀態的鍵值，即上下文的詞向量。V是值向量(Value Vector)，為編碼器隱藏層狀態的輸出值。Q是查詢向量(Query Vector)，為解碼器的前一期輸出。所以自注意力機制對應Q、K、V，共有三種權重，利用神經網路優化可以找到三種權重的最佳值，然後，以輸入向量個別成以三種權重，即可以求得Q、K、V向量。

Q、K、V在經過下圖的運算公式(2)即可以得到自注意力矩陣，其中點積運算：Q ×K，計算輸入向量與上下文詞彙的相似度。再來是特徵縮放Q、K維度開根號，通常Q、K維度是64，所以，故 = 8。最後一步是Softmax運算，將上述轉化為機率，找出要重視的上下文詞彙，用Value Vector乘以上述機率，較大值為要重視的上下文詞彙。

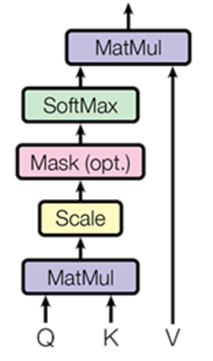


圖3 - 14、自注意力機制運算方式

多標頭的自注意力機制是 (Multi - Head)，通常是八個標頭，如圖3 - 14的機制，經過點內積運算，串聯這八個標頭如下圖3 - 15，多標頭自注意力矩陣公式如下：

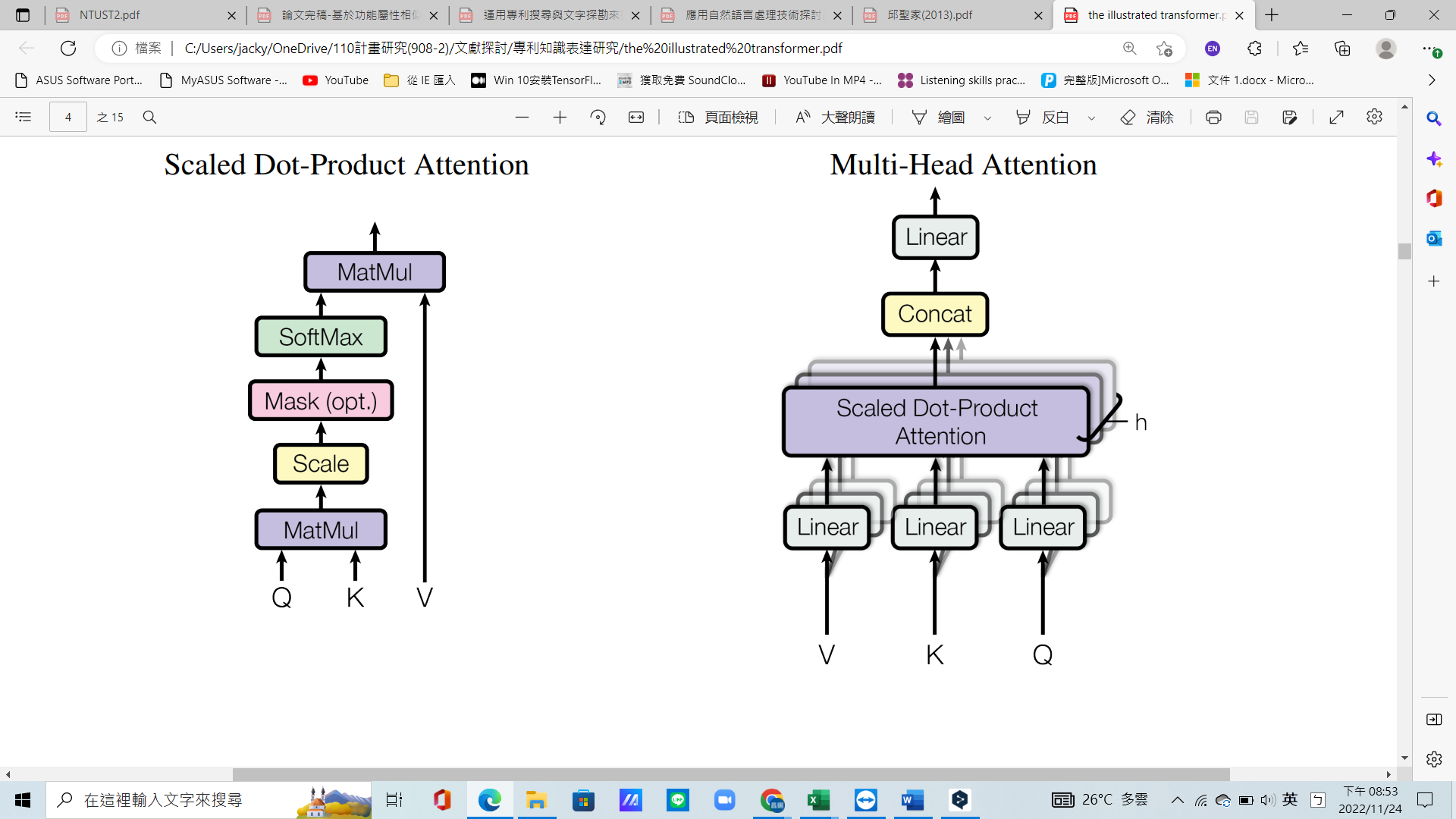


圖3 - 15、多標頭自注意力機制運算方式

**3.5.2****餘弦相似度**

由於Transformer不包含遞歸和卷積，為了使模型能夠利用序列的順序，必須注入一些關於序列中標記的相對位置信息，本研究是計算句子與字詞間的語義相似度，主要是透過餘弦相似度的計算，計算專利文本的句子與功能屬性之間的相似度，基於餘弦相似度計算，其中本研究定義為專利文本關鍵字、為功能或屬性關鍵字，由下方公式表示：

1. **研究結果**

本章節會對上一章的第二、三、四節分析結果進一步說明仔細，畢竟上一章節只是提供概念的可行性為主，實際執行結果並沒有詳細描述。再者，也會部份說明Dov2Vec虛擬碼的內容，把資料集全部用表格的方式呈現。最後，當然也是本研究主要的核心，也就是上述的BERT微調訓練模型所執行的相關結果，會呈現出虛擬碼並加以說明。

**第一節 Doc2Vec協助建立相關性分析結果**

透過研究方法的Doc2Vec將句子轉換為向量並判斷相似度，為發明原則與功能屬性以及專利文本建構分類基礎，取代人工判斷方式，再利用餘弦相似度計算句子與單字或是句子與句子之間的關係，執行BERT訓練之前，把所需建立的前置作業整理好。

**4.1.1驗證發明原則與功能屬性資料集彙整**

首先，下方虛擬碼是用來分析發明原則與功能屬性字詞庫之間的相似度。以下將說明本節虛擬碼的主要內容，這段虛擬碼主要是對語料進行預處理，並讀取停用詞。首先，它會載入numpy、nltk、gensim…等函式庫，nltk是自然語言處理的常用函式庫，gensim是處理語料的工具，接著它會使用nltk函式庫載入停用詞(stopwords)，再從gensim載入Word2Vec和Doc2Vec，這兩個是用來訓練文本向量的工具，最後，它會載入sklearn的cosine similarity，這是用來計算文本之間的相似度的函式。

讀取內文之後，並將其內容存入變數語料庫(corpus)，然後它會設定一個參數MAX WORDS A LINE，用來表示每行最多有幾個字，最後，它會使用string函式庫讀取標點符號，並將停用詞、標點符號和換行符號合併成一個集合，存入變數stopword list如圖4 - 1。

*Import the required libraries "numpy", "nltk", "gensim", "Word2Vec", "Doc2Vec", "TaggedDocument", and "cosine\_similarity".*

*Open a file called "/content/drive/MyDrive/Colab Notebooks/發明原則內文.csv", read its contents and store them in a list called "corpus".*

*Print "corpus".*

*Set a constant "MAX\_WORDS\_A\_LINE" to 500.*

*Print the punctuation symbols.*

*Store the stopwords in English and the punctuation symbols in a set called "stopword\_list".*

圖4 - 1、Doc2Vec虛擬碼說明

下方虛擬碼的目的是對語料進行分詞(Tokenization)，並訓練Doc2Vec模型。首先，定義一個函式tokenize，用來對文本進行分詞，函式的參數有文本、停用詞集合和最大字數(max len)並使用gensim的simple preprocess對文本進行分詞，再使用迴圈將分詞後的結果與停用詞比對，留下不在停用詞集合內的字詞。接著，它會使用一個迴圈對變數語料庫(corpus)中的每行文本呼叫tokenize函式，將分詞後的結果存入document tokens，然後，它會使用Gensim的Tagged Document將documen tokens中的每個元素轉換Tagged Document格式，存入tagged corpus中，至於Tagged Document的作用是將文本和標記(例如：文本的ID)對應起來，最後它會使用Doc2Vec建立一個模型，並使用train函式對模型進行訓練，其中的參數vector size表示文本向量的維度，參數epochs表示訓練輪數，訓練完成後，模型就可以用來計算文本之間的相似度，如圖4 - 2。

*Define a function "tokenize" that takes in three parameters "text", "stopwords", and "max\_len".*

*In the function, use gensim's simple\_preprocess to preprocess the "text", limiting the length to "max\_len".*

*Return a list of tokens from the preprocessed text, excluding those in the "stopwords" list.*

*Initialize an empty list "document\_tokens".*

*For each line in "corpus", use the "tokenize" function to preprocess the line and append the resulting list of tokens to "document\_tokens".*

*Tag each document in "document\_tokens" with its index and store the result in a list "tagged\_corpus".*

*Train a "Doc2Vec" model on "tagged\_corpus" with vector size equal to "MAX\_WORDS\_A\_LINE" and 200 epochs, and store the model in a variable "model\_d2v".*

*Train the "model\_d2v" model on "tagged\_corpus" using the "train" method, providing the number of examples in the corpus and the number of epochs.*

圖4 - 2、Doc2Vec虛擬碼說明

下方虛擬碼主要是對模型進行測試，輸入一個文本，然後計算該輸入的文本與其他內文的文本之間的相似度，並選出相似度最高的前十個內文文本。首先，使用迴圈對document tokens中的每個元素呼叫model d2v的infer vector函式，將分詞後的文本轉換為向量，並存入questions中，接著會使用numpy的cosine similarity函式計算輸入的文本和questions中的每個向量之間的相似度然後，它會將相似度從大到小排序，取出前十名，最後使用迴圈對top n中的每個元素輸出相似度和對應的文本，如圖4 - 3。

*Create an empty list named "questions".*

*For each index i in the range of the length of document\_tokens, append the result of calling infer\_vector on the i - th element of document\_tokens using model\_d2v to questions.*

*Convert questions to a numpy array.*

*Create a variable named "text" with a string value.*

*Create a variable named "filtered\_tokens" by passing text and stopword\_list to the tokenize function.*

*Compute cosine similarity between the result of calling infer\_vector on filtered\_tokens and reshaped to (1,- 1) and questions, with dense\_output set to False. Store the result in "similarity".*

*Create a variable "top\_n" that stores the result of sorting the first element of similarity in descending order, reversed, and only selecting the first 10 elements.*

*For each index in top\_n, print the rounded similarity score at the corresponding index of similarity and the corresponding line of corpus.*

圖4 - 3、Doc2Vec虛擬碼說明

上述所說輸入一個文本，然後計算該文本與其他文本之間的相似度，這邊的文本指在本研究套用的是第三章第一節所展示的功能以及屬性的字詞庫如圖3 - 1。透過上方所說明的虛擬碼，將發明原則與功能屬性做相似度匹配，整理之後的完整表格如附錄，一共有40筆發明原則資料集，下方表格4 - 1展示前5筆發明原則所匹配到的功能以及屬性，這邊因為屬性字詞有延伸字詞或是擴增字詞，所以編號的時候以英文字母小寫做排序，在延伸字詞與擴增字詞中，其語義還是會跟原本(a)開頭的字詞有所差別，因為有些字的意思是次要的，還是以它翻譯時的主要意思為主，所以在匹配發明原則相似度時，並不會將相同數字的英文字母屬性都放在同一個發明原則類別裡面。

表4 - 1、40發明原則功能屬性分項列表

|  |  |
| --- | --- |
| 發明原則 | 功能/屬性 |
| 1.分割 | 功能:5、39、47 |
| 屬性:3(b)、8(b)、8(f)、9(e)、23(d)、25(a)、32、37(d)、40、46(a)、60(a)、62(b)、72、73(d)、73(f)、73(i)、74(c)、75(a) |
| 2.萃取 | 功能:14、55 |
| 屬性:2(d)、3(a)、27(d)、29(a)、37(d)、49、54(e)、73(g) |
| 3.局部品質 | 功能:2、4、5、9、10、14、26、37、40、40、46、47、54、55 |
| 屬性:2(a)、2(d)、3(b)、4(b)、4(c)、6、7(b)、7(c)、8(b)、8(f)、9(a)、11(b)、14、15、18(a)、18(b)、18(c)、19、20(a)、23(c)、23(d)、25(c)、27(c)、27(d)、29(a)、29(b)、29(b)、30、37(a)、37(d)、37(g)、40、42、43(f)、47、50(b)、53(b)、54(b)、54(c)、54(e)、58、60(a)、68(a)、68(b)、68(d)、70、71(c)、73(a)、73(c)、73(d)、73(h)、74(c)、75(a)、75(f) |
| 4.非對稱性 | 功能:20、37、40、55 |
| 屬性:1、3(b)、4(b)、4(c)、5(a)、7(c)、11(c)、11(d)、14、18(b)、19、20(a)、23(d)、25(b)、27(d)、40、42、43(a)、60(a)、62(a)、62(b)、70、71(c)、73(a)、73(i)、75(f) |
| 5.組合合併 | 功能:4、14、23、26、27、46 |
| 屬性:3(a)、3(a)、15、17(b)、29(b)、43(f)、55(b)、59(b)、63(a)、73(j)、75(a) |

**4.1.2驗證發明原則與專利資料集彙整**

驗證此次實驗數據庫是專利摘要的內文全文CSV檔如圖4 - 4，一共是5,489筆專利資料，讓Doc2vec讀取專利ID與摘要文本進行訓練，因為內文資料量較大需要較久時間進行模型訓練，相對的，輸入的文本是使用TRIZ網站上的40發明原則語句說明進行訓練如圖4 - 5，用訓練後的文檔向量進行語義相似度計算，再使用發明原則語句去做相似度比對。

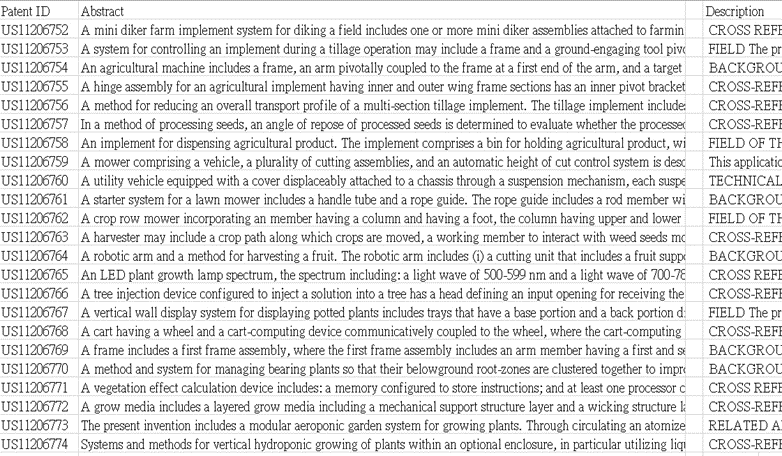


圖4 - 4、專利本地數據庫

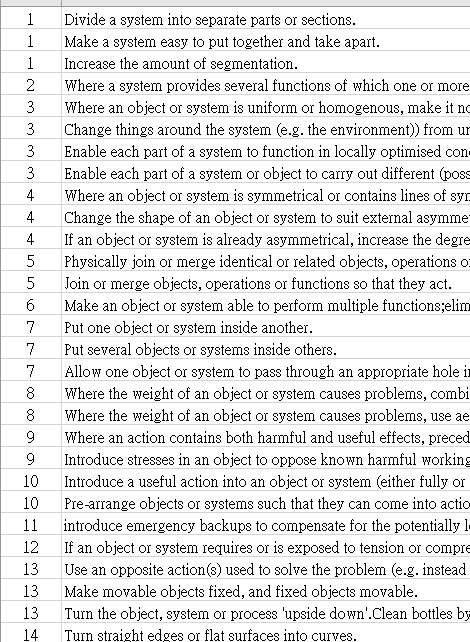


圖4 - 5、發明原則與專利相似度比對

統整後的專利資料集，進行相似度比對就可以知道哪筆專利資料與輸入的發明原則有語義相似之處，接下來進行專利標籤，因摘要內容過於攏長，這邊只顯示專利編號，一共有213筆標籤好發明原則的專利，這邊節錄前20筆如表4 - 2。

表4 - 2、專利少量標籤

|  |  |
| --- | --- |
| Patent ID | 發明原則 |
| US11206756 | 36.Phase transitions |
| US11206782 | 7.Nested doll |
| US11206783 | 7.Nested doll |
| US11206788 | 7.Nested doll |
| US11206794 | 17.Another dimension |
| US11206816 | 2.Taking out |
| US11206839 | 23.Feedback |
| US11206858 | 35.Parameter changes |
| US11206871 | 20.Continuity of useful action |
| US11206876 | 16.Partial or excessive actions |
| US11206892 | 2.Taking out |
| US11206893 | 9.Preliminary anti - action |
| US11206900 | 35.Parameter changes |
| US11206929 | 27.Cheap short - living objects |
| US11206936 | 14.Spheroidality - Curvature |
| US11206949 | 18.Mechanical vibration |
| US11206952 | 27.Cheap short - living objects |
| US11206977 | 3.Local quality |
| US11207056 | 37.Thermal expansion |
| US11207111 | 23.Feedback |

**4.1.3驗證功能屬性與專利資料集彙整**

為了驗證專利文本摘要與功能屬性相似性，運用Doc2vec客觀比較專利摘要與功能屬性相似之處，此次實驗數據庫是專利摘要的內文全文CSV檔如圖4 - 4，讓Doc2vec讀取專利ID與摘要文本進行訓練再輸入文本，這裡輸入的文本為功能或屬性的關鍵字詞庫去做相似度比對，第三章第四節有展示屬性與專利文本的相似度範例如圖3 - 11。下方表4 - 3為統整過後的資料集，一共有213筆標籤好功能屬性的專利，這邊節錄前20筆。

表4 - 3、功能屬性少量標籤

|  |  |  |
| --- | --- | --- |
| Patent ID | 功能 | 屬性 |
| US11206756 | 5.Break down | 46(a).operability |
| US11206782 | 15.Decompose/Separate | 73(c).electric magnetic |
| US11206783 | 15.Decompose/Separate | 73(b).differential potential |
| US11206788 | 16.Protect | 71(a).variety |
| US11206794 | 49.Resist/Retard | 73(h).potential |
| US11206816 | 14.Corrode | 75(d).weighting  factor |
| US11206839 | 8.Change Phase | 23(c).length time |
| US11206858 | 33.Irradiate | 4(a).amount substance |
| US11206871 | 11.Condense | 18(e).electrical path |
| US11206876 | 31.Hold | 59(d).pattern |
| US11206892 | 18.Destroy | 38(c).level mechanization |
| US11206893 | 8.Change Phase | 59(b).form |
| US11206900 | 7.Change | 21(b).measuring |
| US11206929 | 35.Lift | 38(a).level automation |
| US11206936 | 32.Ionize | 5(a).aesthetic |
| US11206949 | 30.Heat | 40.magnetic field |
| US11206952 | 22.Drop | 59(c).configuration |
| US11206977 | 37.Measure | 2(d).accurateness |
| US11207056 | 41.Orient | 60(b).ultrasound |
| US11207111 | 39.Mix | 49.porosity |

**第二節 BERT專利文本與發明原則分析結果**

上一節的專利資料集進行整理，透過Doc2Vec轉換成句子向量進行相似度比對就可以知道哪筆專利資料與輸入的發明原則有語義相似之處，接下來進行專利標籤，一共有213筆標籤好發明原則的專利，這邊節錄前20筆，如圖4 - 6。這個部分屬於第三章第四節所提到的少量標籤，透過少量標籤讓BERT的自注意力機制讀取發明原則標籤與專利摘要內文之間的相似處，進行訓練進而形成相關專利知識結構。

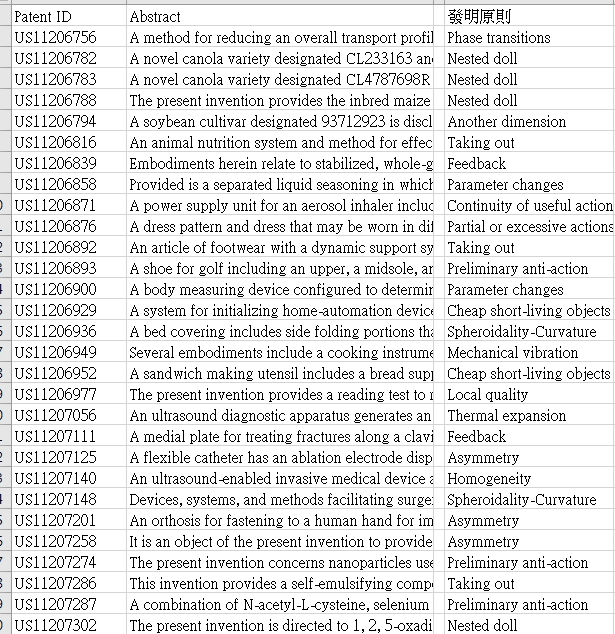


圖4 - 6、發明原則專利標籤

**4.2.1分類效果驗證**

透過上面少量標籤，從5,489筆資料選取213筆專利做發明原則標籤，讓BERT讀取標籤資料與專利摘要內文之間的相似度，接下來進行下游任務訓練，BERT會調用訓練好的模型，來判斷上方真實標籤213筆資料中隨機取20筆樣本出來做預測值，計算模型標籤的準確率與一致性，如表4 - 4。

表4 - 4、發明原則分類結果

|  |  |  |
| --- | --- | --- |
| Patent ID | 發明原則 | 本實驗分類結果 |
| US11206876 | 16 | 35、16、9、1、25 |
| US11206949 | 18 | 33、39、18、7、35、16 |
| US11207056 | 37 | 21、38、2、13、37、6 |
| US11207307 | 29 | 27、40、29、12、23、30 |
| US11207547 | 23 | 26、12、23、33、39、24、 |
| US11207756 | 28 | 15、23、18、28、35、19 |
| US11207876 | 40 | 11、27、40、20、34、12 |
| US11208374 | 9 | 19、31、9、1、21、14 |
| US11208539 | 21 | 1、25、21、14、38、6 |
| US11208565 | 11 | 8、4、11、27、40、20 |
| US11208641 | 27 | 17、4、11、27、22、12、23 |
| US11208777 | 8 | 21、14、38、2、13、37、6 |
| US11209728 | 27 | 4、22、27、32、20、15 |
| US11209757 | 20 | 8、4、11、27、40、20 |
| US11210849 | 33 | 40、12、33、39、24、3 |
| US11211283 | 40 | 39、24、3、5、7、35 |
| US11211585 | 30 | 27、20、26、23、33 |
| US11211887 | 11 | 8、11、27、32、40、20 |
| US11212380 | 10 | 11、27、40、26、15、33 |
| US11212684 | 21 | 1、3、21、25、33、35 |

本研究以嚴格準確率作為評估指標，其算法則是將正確預測的樣本數除以總樣本數進行計算，所以根據以上敘述，這邊假設分子為符合預測發明原則的專利數量，分母為總專利數，公式如下：

由上方發明原則標籤結果得知，該訓練模型會標出有可能是這篇專利號的標籤，大致會列出6筆左右的標籤數量，其應該是大範圍的搜索多標頭注意力機制，找出文字的邏輯性與語境，讓標頭在不同的表示子空間學會關注不同位置的詞彙向量，至於其準確率定義的話為與原始設定的標籤相同，有提及即可算準確，結果20筆總專利數中有16筆符合預測發明原則的專利數量，所以以上標籤的準確率為80%。

**第三節 BERT專利文本與功能屬性參數辨識結果**

透過上述少量標籤，從5,489筆資料選取213筆專利做功能屬性標籤，這邊節錄前20筆資料如圖4 - 7，讓BERT讀取標籤資料與專利摘要內文之間的相似度，透過下游任務訓練，進而形成相關專利知識結構，為了方便瀏覽，本節把功能跟屬性分成小段落做對照。



圖4 - 7、功能屬性專利標籤

**4.3.1功能參數分類效果驗證**

下方表格為專利模型訓練後，一個測試的結果，其方法為從原本標籤好的213筆專利資料，隨機取出20筆做預測，所以可以看出原本所標籤好的功能參數與透過訓練好的模型所標籤的功能參數，藉此結果來證明此模型是否與原本標籤有一致性，如表4 - 5。

表4 - 5、功能參數分類結果

|  |  |  |
| --- | --- | --- |
| Patent ID | 功能參數 | 本實驗分類結果 |
| US11206876 | 31 | 11、31、32 |
| US11206949 | 30 | 26、30、46 |
| US11207056 | 41 | 26、37、41、45 |
| US11207307 | 16 | 10、16、21 |
| US11207547 | 42 | 9、11、42、45 |
| US11207756 | 3 | 3、7、10、13 |
| US11207876 | 54 | 5、7、8、54 |
| US11208374 | 45 | 15、18、20、45 |
| US11208539 | 4 | 4、24、27、39 |
| US11208565 | 53 | 39、42、46、53 |
| US11208641 | 37 | 34、37、39 |
| US11208777 | 4 | 15、17、45、47 |
| US11209728 | 4 | 4、16、18 |
| US11209757 | 11 | 11、13、28、31 |
| US11210849 | 48 | 37、39、48 |
| US11211283 | 11 | 25、26、30、34 |
| US11211585 | 24 | 18、20、22、23 |
| US11211887 | 41 | 33、35、41 |
| US11212380 | 6 | 17、20、39、44 |
| US11212684 | 10 | 3、7、10、34、36、37 |

由上方功能標籤結果得知，該訓練模型會標出有可能是這篇專利號的功能標籤，大致會列出3到4筆左右的標籤數量，其應該是大範圍的搜索多標頭注意力機制，找出文字的邏輯性與語境，讓標頭在不同的表示子空間學會關注不同位置的詞彙向量，至於其準確率定義的話為與原始設定的標籤相同，有提及即可算準確，結果20筆總專利數中中有16筆符合預測功能參數的專利數量，所以以上標籤的準確率為80%。

**4.3.2屬性參數分類效果驗證**

下方表格看出原本所標籤好的屬性參數與透過調用訓練好的模型所標籤的屬性參數，藉此結果來證明此模型是否與原本標籤有一致性。這邊只顯示專利的編號，以利屬性編號對照如表4 - 6。

表4 - 6、屬性參數分類結果

|  |  |  |
| --- | --- | --- |
| Patent ID | 屬性參數 | 本實驗分類結果 |
| US11206876 | 59(d) | 36(a)、8(h)、59(b)、59(d) |
| US11206949 | 40 | 21(b)、40、41、43(a) |
| US11207056 | 60(b) | 60(a)、60(b)、62(b)、71(c) |
| US11207307 | 55(a) | 8(b)、43(g)、55(a) |
| US11207547 | 60(b) | 52、60(b)、62(b) |
| US11207756 | 23(d) | 23(d)、28(b)、38(b)、62(b) |
| US11207876 | 8(a) | 8(a)、22(b)、62(b)、74(d) |
| US11208374 | 59(b) | 5(a)、20(a)、59(b)、63(a) |
| US11208539 | 15 | 15、24(b)、25(a)、59(a) |
| US11208565 | 56 | 23(c)、32、56、74(b) |
| US11208641 | 75(b) | 37(f)、45、55(c)、75(b) |
| US11208777 | 54(a) | 37(b)、38(c)、44 |
| US11209728 | 37(a) | 20(e)、35(a)、63(b) |
| US11209757 | 11(a) | 5(a)、11(a)、11(b)、20(b) |
| US11210849 | 53(a) | 2(b)、20(c)、25(a)、43(d) |
| US11211283 | 74(c) | 71(b)、11(d)、18(e) 、73(c) |
| US11211585 | 8(d) | 31、37(g)、33、43(d) |
| US11211887 | 61(a) | 3(a)、3(c)、18(b)、61(a) |
| US11212380 | 73(a) | 8(e)、24(a)、24(b)、36(b) |
| US11212684 | 41 | 16、41、59(c)、62(b)、20(a) |

由上方屬性標籤結果得知，該訓練模型會標出有可能是這篇專利號的屬性標籤，大致會列出3到4筆左右的標籤數量，其原理為上述所提到的多標頭注意力機制，至於其準確率定義的話為與原始設定的標籤相同，有提及即可算準確，結果20筆總專利數中有14筆符合預測屬性參數的專利數量所以以上標籤的準確率為70%。

**第四節 專利標籤與產品設計之模式分析**

本節透過BERT下游任務訓練，進行多標籤專利分類，給予模型未標籤過的專利內文，並由模型判斷此專利的發明原則、功能、屬性標籤，為了驗證此模型的可行性，會使用人工瀏覽專利的摘要內文並進行判斷，確認模型所標籤的參數是否符合萃智邏輯，。表格內的專利摘要內容放在表格的下方，為了確認人工判讀的合理性，本研究把相關段落內容標是為黃色部分，方便讀者瀏覽。

表4 - 7、驗證發明原則標籤

|  |  |  |
| --- | --- | --- |
| 未標籤Patent ID | 發明原則參數 | 人工判讀 |
| US11208115 | 5、23、27、35 | 23.回饋(Feedback) |
| US11208118 | 5、23、27、35、37 | 23.回饋(Feedback) |
| US11208120 | 5、8、11、26、35 | 5.合併(merging) |
| US11208121 | 2、5、35、25 | 5.合併(merging) |
| US11208132 | 4、5、11、26、37 | 4.非對稱性(asymmetry)、  11.事先預防  (beforehand cushioning) |

表4 - 8、驗證功能參數標籤

|  |  |  |
| --- | --- | --- |
| 未標籤Patent ID | 功能參數 | 人工判讀 |
| US11208115 | 8、19、48、52 | 52.擴散(Spread) |
| US11208118 | 8、29、48、52 | 8.改變相位(Change Phase) |
| US11208120 | 8、16、19、48 | 8.改變相位(Change Phase) |
| US11208121 | 8、29、34、48 | 34.結合(Join) |
| US11208132 | 18、19、39 | 39.混合(Mix) |

表4 - 9、驗證屬性參數標籤

|  |  |  |
| --- | --- | --- |
| 未標籤Patent ID | 屬性參數 | 人工判讀 |
| US11208115 | 9(a)、53(b)、62(a)、68(a) | 53(b).精密度(preciseness)、62(a).物體穩定性(stability object) |
| US11208118 | 53(b)、59(c)、62(a)、68(a) | 53(b).精密度(preciseness)、68(a).時間(time) |
| US11208120 | 9(a)、28(d)、59(c)、68(a) | 68(a).時間(time) |
| US11208121 | 59(c)、68(a) | 59(c).形狀(configuration) |
| US11208132 | 27(a)、28(d)、59(c)、43(e) | 59(c).形狀(configuration) |

專利US11208115原文：*A method of assisting an autonomous vehicle includes obtaining first surrounding area information of the vehicle when the vehicle is located at a first distance from a monitored area disposed ahead of the vehicle、 providing a first control command for controlling the vehicle to operate in a first operating mode、 by using the first surrounding area information、 obtaining second surrounding area information of the vehicle when the vehicle has driven toward the monitored area and is located at a second distance less than the first distance from the monitored area、 and providing a second control command for controlling the vehicle to operate in a second operating mode、 by using the second surrounding area information.*

透過上述專利摘要描述，可以透過黃色標識區域看出這篇專利主要透過獲得車輛區域的周邊訊息，來控制車輛的操作模式，這個與發明原則的編號「23.回饋：導入回饋以改善製造或作用」的概念相似，所以判定為編號23。然而，該訓練模型它的功能與屬性是透過內文的描述給予其類別的分類，透過人工判讀方式其發明是由周邊區域訊息來判定車輛操作模式，所以所屬類別是功能52.擴散(Spread)與屬性53(b).精密度(preciseness)、62(a).物體穩定性(stability object)。

專利US11208118原文：*A travel control device for controlling travel of a vehicle、comprises a device that acquires external information of a vehicle; a travel control unit configured to control travel of the vehicle using an acquisition result by the device; a diagnosis start unit configured to start a diagnosis process of the device; and a restriction unit configured to restrict a function of travel control of the vehicle after the diagnosis process of the device is started by the diagnosis start unit according to a state of travel control of the vehicle before the diagnosis process of the device is started.*

上述專利摘要描述，黃色標識區域顯示此控制裝置需要獲取車輛外部信息，以控制車輛的行駛，透或獲取外部信息以及行駛控制單元可以看出與發明原則的編號「23.回饋：如果已使用回饋機制，使能適應作業條件的變化」的概念相似，所以判定為編號23。然而，該訓練模型它的功能與屬性是透過內文的描述給予其類別的分類，透過人工判讀方式，該裝置的診斷過程啟動前車輛的行駛控制狀態限制與功能8.改變相位(Change Phase)與屬性53(b).精密度(preciseness)、68(a).時間(time)有關連。

專利US11208120原文：*Systems and methods for a mobile crane are provided.In particular, systems and methods are provided for a mobile crane that is designed to enable easy transportation on a railcar to and from a job site. The mobile crane facilitates quick changing between a transport configuration and a work configuration to enable increased productivity. The mobile crane is configured to be modular and thereby be configured to include various lifting and object manipulating features, as required by a specific application.*

上述專利摘要描述，黃色標識區域顯示此裝置该移動式起重機便於在運輸配置和工作配置之間快速轉換可以看出與發明原則的編號「5.合併：合併物體、作業或功能使其在時間上能一起作用」的概念相似，所以判定為編號5。然而，該訓練模型它的功能與屬性是透過內文的描述給予其類別的分類，透過人工判讀方式，該起重機發明設計成運輸配置和工作配置之間轉換以改變生產率與功能8.改變相位(Change Phase)與屬性68(a).時間(time)有關連。

專利US11208121原文：*A modern scenic passenger railcar system is disclosed. The railcar system incorporates design and safety features of modern freightcars and modern passenger scenic railway cars. The railcar system incorporates a center beam with a square cross section as well as specialized crash posts with its frame. This results in enhanced safety and crashworthiness. The railcar system also incorporates cross members for absorbing and distributing shock and mechanical stress on the railcar system during use, by withstanding torsion and shear forces on the frame. This results in improved mechanical integrity of the frame and the railcar system overall.*

上述專利摘要描述，黃色標識區域顯示該軌道車系統結合了現代貨運車和現代客運景區鐵路車的設計和安全特徵可以看出與發明原則的編號「5.合併：將相同或相關的物體、作業或功能實體連接或合併」的概念相似，所以判定為編號5。然而，該訓練模型它的功能與屬性是透過內文的描述給予其類別的分類，透過人工判讀方式，該軌道車系統將一個具有方形截面的中心樑以及專門的防撞柱與它的框架結合起來與功能34.結合(Join)與屬性59(c).形狀(configuration)有關連。

專利US11208132原文：*The invention is a wheeled folding tray cart and method. The wheeled folding tray cart comprises a collapsible wheeled framework pivotally supporting a set of split shelves, each split shelf comprises a pair of half shelves. The wheeled folding tray cart moves between an open operating position wherein the set of split shelves are held in spaced - apart, stacked, horizontal orientation within the collapsible wheeled framework and a closed storage position wherein the collapsible wheeled framework is collapsed in manner to move each half shelf pair into an accordion - like, inverted shaped or accordion like configuration. An extension coiled spring attached to the collapsible wheeled framework assists in the movement of the wheeled folding tray cart between open operating position and the closed storage position. A sheath encloses at least a portion of the spring to provide resistance to the spring's movement to limit potential operator exposure to pinch injuries.*

上述專利摘要描述，黃色標識區域顯示带輪子的摺疊式托盤車在儲存上，架子對被摺疊成手風琴狀和一個護套限制操作員可能受到的夾傷可以看出與發明原則的編號「4.非對稱性：使用非對稱的形式取代對稱形式」「11.事先預防：事先準備緊急的方法以補救物體潛在的低可靠度」的概念相似，所以判定為編號4、11。然而，該訓練模型它的功能與屬性是透過內文的描述給予其類別的分類，透過人工判讀方式，連接到可摺疊輪式框架上的拉伸卷簧協助輪式摺疊托盤車在打開的操作位置和關閉的存儲位置之間移動與功能39.混合(Mix)與屬性59(c).形狀(configuration)有關連。

1. **結論與建議**

在這不斷求變求新的時代，人類追求感官、物質生活活已經昇華到另一個層次，就是所謂的AI智慧，它正在往人類的思維模式去做感善，運用AI協助人類在處理事務時的效率。在產品開發已經不在需要由簡單的調整參數或是耗費大量時間瀏覽複雜的文字資料來解決設計需求，本研究利用專利的摘要內文，萃取出可能有用的參考價值，並透過分析專利的發明原則與功能屬性的相似之處，來提供產品大方向探索的實際導向，並將其傳達给產品設計者作為創新設計的發想參考

**第一節 結論**

本研究提出一套基於TRIZ的知識結構，使得專利內容能夠轉化成知識應用的流程架構，希望能將專利文本藉由自然語言處利技術來萃取技術特徵。經由文獻研究，BERT模型導入專利資料進行模型訓練，再通過Doc2Vec建立發明原則、56個功能與75個屬性的關聯性，以及少量運用Doc2Vec進行相似度比對的標籤，將有助於得到知識結構的组合。本研究訓練出模型讓未標籤的專利擁有發明原則的知識結構以及功能屬性，讓創新產業開發人員能迅速了解欲選擇專利的知識結構基礎，並在尋找產品適合發明原則大方向改善時，使用本研究所統整好的發明原則分類表格，找出細項的實際改善方式，以利創新產業開發研究人員縮短開發產品時間並減少人力時間成本。本研究也透過隨機取出5篇的未標籤專利摘要內文，進行案例驗證說明，確認人工判讀的結果與訓練後的模型所標籤的專利一致，證明此研究如未來應用在創新產業的可行性。

本研究主要有三項貢獻:

1. 本研究提出利用(邱聖家, 2013)所建立好的功能屬性關鍵字詞庫做延伸，作為識別專利摘要文本知識內涵，可以透過Doc2Vec語義相似度的計算來得知專利文本與功能屬性的關聯性。
2. 本研究建構出以發明原則結構為基礎，建構發明原則功能屬性分項列表，藉由Doc2Vec語義相似度的分析來識別與發明原則與功能屬性之關聯性
3. 本研究微調BERT訓練出以發明原則定義結構為基礎，以及透過功能屬性判斷標準來識別專利知識結構的模型。

**第二節 研究限制**

本研究使用非監督式機器學習做標籤，用監督式機器學習做訓練，因此研究結果需要人工審核驗證其可行性。由於專利標籤資料量的不足，所以導致在驗證過程中屬性的字詞標籤訓練成果不佳，準確率下降至7成，目前判定可能為專利標籤資料量不足所導致，當然也有可能為在初始化標籤時，需要用到標籤定義的方式去做相似度分析，本研究在建立發明原則與專利標籤時，是用定義去做少量標籤，但是在功能屬性標籤時，並沒有用其定義的方式去做標籤，而是直接讓Doc2Vec去判斷其標籤本身的單詞向量去與文本做相似度比對，所以如果能在初始化時用功能屬性字詞庫的定義去做相似度比對，或許模型結果準確率會更高。

在本身模型訓練時，如果使用Jupyter跑資料訓練，常常因為內存不足而導致無法順利輸出模型，如果使用Colab上的RAM為24GB的Colab環境，能夠處理更大的和更具記憶體密集型的工作負載或同時執行多個任務而不需要面對內存限制。

**第三節 未來研究與建議**

關於本研究雖然有訓練出能夠給予專利摘要發明原則以及功能屬性標籤，但是還是有許多不足之處，未來可以由此作為研究之方向:

1. 自然語言處理

目前本研究在專利是使用BERT自然語言處理套件來完成發明原則功能屬性判斷分析，有時專利文字數量龐大會導致模型資料過於龐大無法訓練，運用摘要雖然可以擷取知識結構，但是相對於專利描述來說，專利描述或許更加完整。在執行本研究時，Chat - GPT、DALL E 2人工智慧系統剛開發出來，DALL E 2它可以從自然語言的描述中創建逼真的圖像和藝術；Chat - GPT則是從自然語言的文本描述中生成語言文字檔，或許未來研究方向能夠圖文並茂的說明專利知識結構，這部分可深入探討。

1. 功能屬性關鍵字詞庫

本研究所建立的屬性結構關鍵詞詞庫是基於屬性同義詞擴充來實現的，使用多個詞語來表示該屬性的定義。但是，功能詞沒有使用同義詞擴充來實現，如果只使用單一該意義的詞語識別範圍會較為不足，仍有很大的改進空間，或許未來可以整理出功能詞的延伸詞庫，並在初始化時用功能加屬性詞庫的定義去做相似度對比，提高模型訓練準確度。

**參考文獻**

**中文部分：**

* 吳欣玲(2015b)，經濟部智慧財產局改善專利積案之政策分析，國立臺灣大學, 台北市，Retrieved from <https://hdl.handle.net/11296/c7gmpy>。
* 呂宗興(2012)，萃智趨勢分析與應用之研究，未出版之碩士論文，國立清華大學工業工程與工程管理學系，新竹市。
* 林敬勛(2022)，利用GPT - 3產生表單內容應用於自動化網頁填寫代理人，國立臺北科技大學, 台北市，Retrieved from <https://hdl.handle.net/11296/4x3w65>。
* 林瑞庭(2011)，萃智(TRIZ)創新高關聯專利檢索與網路平台建置—以中華民國專利與美國專利為例，未出版之碩士論文，國立臺北科技大學工業工程與管理研究所，台北市，
* 邱聖家(2013)，使用相似性指標辨識萃智解答模型:以相關趨勢辨識為例，未出版之碩士論文，國立清華大學工業工程與工程管理學系，新竹市。
* 科技產業資訊室(2010, 2010/01/07)，工研院成立專利聯盟提升台灣產業競爭力，from <https://iknow.stpi.narl.org.tw/Post/Read.aspx?PostID=4835>。
* 范宇碩(2019)，專利跨域再生的系統化方法，國立清華大學。
* 孫健晟(2020)，PyTorch機械深度學習架構應用之研究，國立雲林科技大學，雲林縣，Retrieved from <https://hdl.handle.net/11296/4k8q4f>。
* 徐楷崴(2022)，基於BERT之情感分析數據標記系統，淡江大學，新北市，Retrieved from <https://hdl.handle.net/11296/2vy2sa>。
* 張家瑋(2012)，英語短文語意相似度評估演算法，國立成功大學，台南市，Retrieved from <https://hdl.handle.net/11296/786fvd>。
* 張易筠(2022)，應用BERT語言模型於顧客評論之多面向情緒分析，國立高雄科技大學，高雄市，Retrieved from <https://hdl.handle.net/11296/4v7q4a>。
* 莊政翰(2007)，運用類神經網路與貝氏網路結合TF - IDF與Entropy特徵於本體論之自動建構，朝陽科技大學，台中市，Retrieved from <https://hdl.handle.net/11296/3zh88p>。
* 陳宛琳(2014)，結合本體論與語意相似程度對文件萃取關鍵字，中原大學，桃園縣，Retrieved from <https://hdl.handle.net/11296/kzy684>。
* 陳達仁(2004)，專利檢索，智慧財產權月刊，70，5 - 11。
* 葉子諒(2017b)，電腦輔助專利創新性評估與檢索系統，國立清華大學，新竹市，Retrieved from <https://hdl.handle.net/11296/en4qa9>。
* 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。
* 鄧乃誠(2015)，基於功能屬性相似性具優先序的萃智趨勢解答辨識，未出版之碩士論文，國立清華大學工業工程與工程管理學系，新竹市。
* 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。

**外文部分：**

* Akay, H., & Kim, S. - G. (2021). Extracting functional requirements from design documentation using machine learning. Procedia CIRP, 100, 31 - 36. doi: <https://doi.org/10.1016/j.procir.2021.05.005>
* Choi, J., & Yoon, J. (2022). Measuring knowledge exploration distance at the patent level: Application of network embedding and citation analysis. Journal of Informetrics, 16(2), 101286. doi: <https://doi.org/10.1016/j.joi.2022.101286>
* Catelli, R., Fujita, H., De Pietro, G., & Esposito, M. (2022). Deceptive reviews and sentiment polarity: Effective link by exploiting BERT. Expert Systems with Applications, 209, 118290. doi: <https://doi.org/10.1016/j.eswa.2022.118290>
* Cong, H., & Tong, L. H. (2008). Grouping of TRIZ Inventive Principles to facilitate automatic patent classification. Expert Systems with Applications, 34(1), 788 - 795. doi: <https://doi.org/10.1016/j.eswa.2006.10.015>
* Devlin, J., Chang, M. - W., Lee, K., & Toutanova, K. (2019). BERT: Pre - training of Deep Bidirectional Transformers for Language Understanding. ArXiv, abs/1810.04805.
* Dourson, S. (2004). The 40 inventive principles of TRIZ applied to finance. The TRIZ journal, 1(1), 1 - 23.
* Gey, F., Buckland, M., Chen, A., & Larson, R. (2001). Entry vocabulary - a technology to enhance digital search. Paper presented at the Proceedings of the first international conference on Human language technology research.
* Guven, Z. A., & Unalir, M. O. (2022). Natural language based analysis of SQuAD: An analytical approach for BERT. Expert Systems with Applications, 195, 116592.
* Hyun, J. S., & Park, C. J. (2016). Compromise: An Alternative Solution Strategy for Contradiction Problems in the Butterfly Model. Procedia CIRP, 39, 103 - 108. doi: <https://doi.org/10.1016/j.procir.2016.01.173>
* Larkey, L. S. (1999). A patent search and classification system. Paper presented at the Proceedings of the fourth ACM conference on Digital libraries.
* Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. Paper presented at the International conference on machine learning.
* Lee, C. - H., Zhao, X., & Lee, Y. - C. (2019). Service quality driven approach for innovative retail service system design and evaluation: A case study. Computers & Industrial Engineering, 135, 275 - 285.
* Li, W. - q., Li, Y., Chen, J., & Hou, C. - y. (2017). Product functional information based automatic patent classification: Method and experimental studies. Information Systems, 67, 71 - 82. doi: <https://doi.org/10.1016/j.is.2017.03.007>
* Li, Z., Tate, D., Lane, C., & Adams, C. (2012). A framework for automatic TRIZ level of invention estimation of patents using natural language processing, knowledge - transfer and patent citation metrics. Computer - Aided Design, 44(10), 987 - 1010. doi: <https://doi.org/10.1016/j.cad.2011.12.006>
* Liu, H., Li, W., & Li, Y. (2021). A new computational method for acquiring effect knowledge to support product innovation. Knowledge - Based Systems, 231, 107410. doi: <https://doi.org/10.1016/j.knosys.2021.107410>
* Liu, L., Li, Y., Xiong, Y., & Cavallucci, D. (2020). A new function - based patent knowledge retrieval tool for conceptual design of innovative products. Computers in Industry, 115, 103154.
* Loh, H. T., He, C., & Shen, L. (2006). Automatic classification of patent documents for TRIZ users. World Patent Information, 28(1), 6 - 13.
* 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.
* Montecchi, T., & Russo, D. (2015). Knowledge based approach for formulating TRIZ contradictions. Procedia Engineering, 131, 451 - 463.
* A multi - granularity knowledge representation and mining method for patent texts. (2022). 6.
* Rajpurkar, P., Jia, R., & Liang, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. arXiv preprint arXiv:1806.03822.
* Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text - conditional image generation with clip latents. arXiv preprint arXiv:2204.06125.
* Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Sutskever, I. (2021). Zero - Shot Text - to - Image Generation. Paper presented at the Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research. <https://proceedings.mlr.press/v139/ramesh21a.html>
* Sarica, S., Luo, J., & Wood, K. L. (2020). TechNet: Technology semantic network based on patent data. Expert Systems with Applications, 142, 112995. doi: <https://doi.org/10.1016/j.eswa.2019.112995>
* Sarkissian, A. (2013). Deciphering innovation: An exploration of USPTO patents granted to Iranian inventors. World Patent Information, 35(4), 313 - 320. doi: <https://doi.org/10.1016/j.wpi.2013.06.005>
* Savransky, S. D. (2000). Engineering of creativity: Introduction to TRIZ methodology of inventive problem solving: CRC press.
* Sheu, D. D., Chiu, M. - C., & Cayard, D. (2020). The 7 pillars of TRIZ philosophies. Computers & Industrial Engineering, 146, 106572. doi: <https://doi.org/10.1016/j.cie.2020.106572>
* Sheu, D. D., & Chiu, S. C. (2017). Prioritized relevant trend identification for problem solving based on quantitative measures. Computers & Industrial Engineering, 107, 327 - 344. doi: <https://doi.org/10.1016/j.cie.2016.03.028>
* Sheu, D. D., & Hong, J. (2018). Prioritized relevant effect identification for problem solving based on similarity measures. Expert Systems with Applications, 100, 211 - 223. doi: <https://doi.org/10.1016/j.eswa.2018.01.032>
* Sheu, D. D., Hong, J., & Ho, C. - L. (2017). New product identification and design through super - system trimming. Computers & Industrial Engineering, 111, 251 - 262.
* Souili, A., Cavallucci, D., Rousselot, F., & Zanni, C. (2015). Starting from Patents to Find Inputs to the Problem Graph Model of IDM - TRIZ. Procedia Engineering, 131, 150 - 161. doi: <https://doi.org/10.1016/j.proeng.2015.12.365>
* Tan, S., Zeng, Y., Chen, B., Bani Milhim, H. K., & Schiffauerova, A. (2012). Environment Based Design Approach to Integrating Enterprise Applications. Journal of Computing and Information Science in Engineering, 12(3). doi: 10.1115/1.4007171
* Terninko, J., Zusman, A., & Zlotin, B. (1998). Systematic innovation: an introduction to TRIZ (theory of inventive problem solving): CRC press.
* Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
* Widad, A., El Habib, B. L., & Ayoub, E. F. (2022). Bert for Question Answering applied on Covid - 19. Procedia Comput Sci, 198, 379 - 384. doi: 10.1016/j.procs.2021.12.257
* Yang, W., Cao, G., Peng, Q., & Sun, Y. (2021). Effective radical innovations using integrated QFD and TRIZ. Computers & Industrial Engineering, 162, 107716. doi: <https://doi.org/10.1016/j.cie.2021.107716>
* Yu, W., Zhu, C., Fang, Y., Yu, D., Wang, S., Xu, Y., Jiang, M. (2022). Dict - BERT: Enhancing Language Model Pre - training with Dictionary. Paper presented at the FINDINGS.
* Zeng, Y. (2004). Environment - based formulation of design problem. Journal of Integrated Design & Process Science, 8.
* Zeng, Y. (2008). Recursive object model (ROM)—Modelling of linguistic information in engineering design. Computers in Industry, 59(6), 612 - 625. doi: <https://doi.org/10.1016/j.compind.2008.03.002>
* Zhang, L., Sun, M., Peng, Y., Zhao, W., Chen, L., & Huang, Y. (2022). How public investment fuels innovation: Clues from government - subsidized USPTO patents. Journal of Informetrics, 16(3), 101313. doi: <https://doi.org/10.1016/j.joi.2022.101313>

**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 specifi ed 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. |

**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"] |

**附錄四、屬性字詞庫**

|  |  |  |
| --- | --- | --- |
| no. | Attribute | Chinese |
| 1 | acceleration | 加速度 |
| 2(a) | accuracy | 準確度 |
| 2(b) | exactness | 準確度 |
| 2(c) | correctness | 準確度 |
| 2(d) | accurateness | 準確度 |
| 3(a) | adaptability | 適應性/彈性 |
| 3(b) | flexibility | 適應性/彈性 |
| 3(c) | suitability | 適應性/彈性 |
| 3(d) | applicability | 適應性/彈性 |
| 4(a) | amount substance | 物質數量 |
| 4(b) | matter | 物質數量 |
| 4(c) | substance type | 物質數量 |
| 5(a) | aesthetic | 美觀 |
| 5(b) | fineness | 美觀 |
| 6 | amplitude | 振幅 |
| 7(a) | angle | 角度 |
| 7(b) | backer | 角度 |
| 7(c) | radian | 角度 |
| 8(a) | area | 面積/表面積 |
| 8(b) | surface area | 面積/表面積 |
| 8(c) | expanse | 面積/表面積 |
| 8(d) | region | 面積/表面積 |
| 8(e) | country | 面積/表面積 |
| 8(f) | process zone | 面積/表面積 |
| 8(g) | zone | 面積/表面積 |
| 8(h) | size | 面積/表面積 |
| 9(a) | brightness | 亮度 |
| 9(b) | luminance | 亮度 |
| 9(c) | intensity | 亮度 |
| 9(d) | lightness | 亮度 |
| 9(e) | luminosity | 亮度 |
| 10 | cleanliness | 潔淨度 |
| 11(a) | color | 顏色 |
| 11(b) | hue | 顏色 |
| 11(c) | pigment | 顏色 |
| 11(d) | dyeware | 顏色 |
| 12 | compatibility | 相容性 |
| 13 | concentration | 濃度 |
| 14 | conductivity electrical | 電導性 |
| 15 | conductivity heat | 熱導性 |
| 16 | controllability | 控制性 |
| 17(a) | cost | 成本 |
| 17(b) | money | 成本 |
| 18(a) | current | 電流 |
| 18(b) | current flow | 電流 |
| 18(c) | current path | 電流 |
| 18(d) | electric current | 電流 |
| 18(e) | electrical path | 電流 |
| 19 | density | 密度 |
| 20(a) | device complexity | 裝置複雜度 |
| 20(b) | equipment complexity | 裝置複雜度 |
| 20(c) | facility complexity | 裝置複雜度 |
| 20(d) | apparatus complexity | 裝置複雜度 |
| 20(e) | installation complexity | 裝置複雜度 |
| 21(a) | difficulty detecting | 偵測/量測困難度 |
| 21(b) | measuring | 偵測/量測困難度 |
| 22(a) | direction | 方向性 |
| 22(b) | orientation | 方向性 |
| 23(a) | duration action | 物體持續時間 |
| 23(b) | action object | 物體持續時間 |
| 23(c) | length time | 物體持續時間 |
| 23(d) | continuance | 物體持續時間 |
| 24(a) | electric field | 電場/電荷強度 |
| 24(b) | electric charge | 電場/電荷強度 |
| 25(a) | energy | 能量 |
| 25(b) | capability | 能量 |
| 25(c) | power source | 能量 |
| 26(a) | fluidity | 流動性 |
| 26(b) | mobility | 流動性 |
| 27(a) | force | 力量 |
| 27(b) | forcefulness | 力量 |
| 27(c) | military force | 力量 |
| 27(d) | military group | 力量 |
| 27(e) | military unit | 力量 |
| 28(a) | frequency | 頻率 |
| 28(b) | frequence | 頻率 |
| 28(c) | oftenness | 頻率 |
| 28(d) | absolute frequency | 頻率 |
| 29(a) | friction | 摩擦力 |
| 29(b) | rubbing | 摩擦力 |
| 30 | function efficiency | 效率 |
| 31 | hardness | 硬度 |
| 32 | harmful factor | 有害因子 |
| 33 | homogeneity | 同質性 |
| 34 | humidity | 濕度 |
| 35(a) | illumination intensity | 照度 |
| 35(b) | illumination | 照度 |
| 36(a) | information | 資訊 |
| 36(b) | data | 資訊 |
| 36(c) | material | 資訊 |
| 37(a) | length | 長度 |
| 37(b) | linear dimension | 長度 |
| 37(c) | displacement | 長度 |
| 37(d) | distance | 長度 |
| 37(e) | size | 長度 |
| 37(f) | width | 長度 |
| 37(g) | height | 長度 |
| 38(a) | level automation | 自動化程度 |
| 38(b) | level mechanisation | 自動化程度 |
| 38(c) | level mechanization | 自動化程度 |
| 39 | lubricity | 潤滑 |
| 40 | magnetic field | 磁場強度 |
| 41 | magnetic flux | 磁通量 |
| 42 | magnetic force | 磁力 |
| 43(a) | maintainability | 維修性 |
| 43(b) | ease repair | 維修性 |
| 43(c) | fix | 維修性 |
| 43(d) | fixing | 維修性 |
| 43(e) | mending | 維修性 |
| 43(f) | repair operation | 維修性 |
| 43(g) | reparation | 維修性 |
| 44 | manufacturability | 可製性 |
| 45 | mass | 質量 |
| 46(a) | operability | 可操作性 |
| 46(b) | functioning | 可操作性 |
| 46(c) | working | 可操作性 |
| 47 | phase state | 相位 |
| 48 | polarity | 極性 |
| 49 | porosity | 多孔性 |
| 50(a) | position | 位置 |
| 50(b) | spatial relation | 位置 |
| 50(c) | place | 位置 |
| 50(d) | distance | 位置 |
| 51 | potential gradient | 電位梯度 |
| 52 | power | 功率 |
| 53(a) | precision | 精密度 |
| 53(b) | preciseness | 精密度 |
| 54(a) | pressure | 壓力/應力 |
| 54(b) | stress | 壓力/應力 |
| 54(c) | pressing | 壓力/應力 |
| 54(d) | mechanical press | 壓力/應力 |
| 54(e) | printing press | 壓力/應力 |
| 55(a) | pressure gradient | 壓力梯度 |
| 55(b) | force per | 壓力梯度 |
| 55(c) | unit area | 壓力梯度 |
| 55(d) | pressure level | 壓力梯度 |
| 56 | productivity | 生產力 |
| 57(a) | purity | 純度 |
| 57(b) | pureness | 純度 |
| 58 | reliability | 可靠度 |
| 59(a) | shape | 形狀 |
| 59(b) | form | 形狀 |
| 59(c) | configuration | 形狀 |
| 59(d) | pattern | 形狀 |
| 59(e) | conformation | 形狀 |
| 60(a) | sound | 聲音 |
| 60(b) | ultrasound | 聲音 |
| 61(a) | speed | 速率 |
| 61(b) | velocity | 速率 |
| 62(a) | stability object | 物體穩定性 |
| 62(b) | constancy object | 物體穩定性 |
| 62(c) | stableness object | 物體穩定性 |
| 63(a) | strength | 強度 |
| 63(b) | intensity level | 強度 |
| 64(a) | surface smoothness | 表面平滑度 |
| 64(b) | surface finish | 表面平滑度 |
| 64(c) | surface smooth | 表面平滑度 |
| 65 | temperature | 溫度 |
| 66 | temperature gradient | 溫度梯度 |
| 67 | thickness | 厚度 |
| 68(a) | time | 時間 |
| 68(b) | period | 時間 |
| 68(c) | while | 時間 |
| 68(d) | lasting | 時間 |
| 68(e) | moment | 時間 |
| 69(a) | transparence | 透明度 |
| 69(b) | translucency | 透明度 |
| 70 | uniformity | 均勻度 |
| 71(a) | variety | 多樣性 |
| 71(b) | diversity | 多樣性 |
| 71(c) | category | 多樣性 |
| 72 | viscosity | 黏滯性 |
| 73(a) | voltage | 電壓 |
| 73(b) | differential potential | 電壓 |
| 73(c) | electric magnetic | 電壓 |
| 73(d) | electric potential | 電壓 |
| 73(e) | electric voltage | 電壓 |
| 73(f) | electromotive force | 電壓 |
| 73(g) | electromotive voltage | 電壓 |
| 73(h) | potential | 電壓 |
| 73(i) | potential difference | 電壓 |
| 73(j) | potential drop | 電壓 |
| 74(a) | volume | 體積/容量 |
| 74(b) | space | 體積/容量 |
| 74(c) | bulk | 體積/容量 |
| 74(d) | cubage | 體積/容量 |
| 74(e) | size | 體積/容量 |
| 75(a) | weight | 重量 |
| 75(b) | weighting | 重量 |
| 75(c) | weighting coefficient | 重量 |
| 75(d) | weighting factor | 重量 |
| 75(e) | weight unit | 重量 |
| 75(f) | system weights | 重量 |