成果報告書格式規範

一、撰寫應包含下列各項：

|  |  |
| --- | --- |
| 標題 | 說明 |
| 封面 | 制式型式，請按型式填寫。 |
| 目錄 | 型式可自行設計，惟應含範例內文字。 |
| 解題成果簡表 | 制式表格，請按表填寫。 |
| 解題執行內容與成果說明 | 惟應含括下列項目：  1. 題目背景說明  2. 解題技術架構及執行方法  3. 成果效益與完成之工作  4. 商業應用價值與創新亮點  5. 結論  6. 附件 |

二、成果報告書編排格式設定：

1. 格式規則：

(1) 版面格式：A4大小直式橫寫（由左至右）製作。本文版面規格邊界上、下各為2公分，左為2.5公分、右為2公分。

(2) 字體設定：中文字型：標楷體、英文字型：Times New Roman；字體：標題16、內文14。頁碼採10號字

(3) 段落設定：行距採單行間距，段落間距與前段與後段距離採0.5行。

2. 頁碼：請以數字標記在每頁下方，以便於查對。

3. 計畫進行各項研究分析、文獻分析、市場調查資料等，應註明資料來源及資料日期。

4. 各項資料編列應注意前後一致，按實編列。

解題成果報告書撰寫說明

解題成果報告書內容至少包括下列各項：

**一、 題目背景說明**

1. 題目說明：說明問題情境、背景、痛點和解題目標等資訊。
2. 構想背景說明：簡述相關問題解決方案之文獻、市調資料回顧，並說明解題構想。

**二、 解題技術架構及執行方法**

所採用之整體技術架構、AI方法、模型及使用情境等相關說明

1. 採用之技術、AI方法與模型
2. 所採用AI方法與模型的優劣與此架構所克服的問題
3. 技術架構圖
4. 使用情境、流程圖等說明應用狀況的資訊(可以Demo檔案代替說明)

**三、 成果效益與完成之工作**

請列述在執行期限內完成之工作項目與帶予出題單位或產業之效益。陳述內容應完整、明確，須包含質化或量化的成果資訊，並請著重說明成果對於出題單位的重要性和效益。

**四、 商業應用價值與創新亮點**

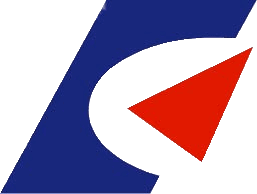
1. 描述此解決方案於其他場景的商轉可行性與應用價值(使用對象、解決的痛點、應用價值與成本、商業模式等)
2. 簡述此解決方案中商業層面、應用層面、技術層面之創新。

**五、 結論**

請總結此次解題為出題單位帶來效益及對自身團隊帶來的成長，同時請評價此次競賽後與出題單位後續合作或是創造其他商業合作的可能性，如有相關文件請列於附件。

**六、 附件**

1. 獎補金核銷表
2. 與出題單位會談之會議紀錄。
3. 解決方案創造之商業合作佐證(如：衍伸之商業合約、政府補助案遞件等證明)



**經濟部工業局**

**AI產業實戰應用人才淬煉計畫**

**110年度解題實證成果報告書**

|  |  |
| --- | --- |
| 解題團隊： | 馬訓冷凝 |
| 題目名稱： | 語意機器人典藏資料搜尋系統 |
| 出題單位： | 文化部 |
| 題目類別： | □ 電腦視覺  ■ 自然語言  □ 數據分析 |

中華民國一百一十年十月十五日

**※申請團隊保證申請文件所列資料及附件均屬正確※**

**※若有偽造不實者或侵權行為，申請團隊須負完全之法律責任※**

**目錄**

* **成果報告摘要表**
* **解題執行內容與成果說明**

一、題目背景說明 13

二、解題技術架構及執行方法 15

三、成果效益與完成之工作 22

四、商轉可行性與成果創新價值 28

五、結論 30

六、附件 30

**AIGO人才解題成果報告摘要表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 團隊名稱 | | 馬訓冷凝 | | | | |
| 題目名稱 | | 語意機器人典藏資料搜尋系統 | | | | |
| 出題單位 | | 文化部 | | | | |
| 題目類別 | | □電腦視覺 ■ 自然語言 □數據分析 | | | | |
| 解題期程 | | 自110年7月2日至10月15日止 | | | | |
| 聯絡人 | | 姓名 | | 黃崇睿 | 電話 | 0921693669 |
| Email | | aa516622@gmail.com | | |
| **團隊任務配置：** AI演算法、資料處理、前後端、PM, etc（參考但不限於） | | | | | | |
| N | 團隊角色 | 姓名/  公司名稱 | | 經歷專長/  公司簡介 | 分工說明 | |
| 1 | 隊長 | 游國忠/  真理大學 | | 自然語言處理/機器視覺/機器學習/深度學習/數據分析/圖像識別/影像處理 | 系統整合、專案管理、企業聯絡 | |
| 2 | 聯絡人 | 黃崇睿/  淡江大學 | | 機器/深度學習 | 知識圖譜、模糊比對、FW搜尋法、句向量、ST搜尋法 | |
| 3 | 隊員 | 郭俞含/  淡江大學 | | 機器/深度學習 | Google Trend, GT搜尋法、系統整合、網頁設計製作 | |
| 4 | 隊員 | 張菘芳/  真理大學 | | 機器/深度學習 | 系統測試 | |
| 計畫執行過程照片與說明（至少四張） | |  | | |  | |
| 由於疫情的關係，本團隊無法前往文化部進行實體的會議，藉由線上會議展示系統。 | | | 透過線上會議與文化部進行簡報報告。 | |
|  | | |  | |
| 與文化部資訊處處長以及副座進行專案的討論 | | | 團隊內部進度討論 | |
| 解題內容  摘要 | | 一、解題背景說明  文化部國家文化記憶庫網站是台灣重要文化典藏資訊來源。該網站製作精美，且有很親民的使用者介面，但文字搜尋機制主要還是依靠關鍵字字串比對。使用關鍵字字串比對最大的問題是，當民眾對於文化資料及專業術語較為陌生，將難以利用精確的關鍵字由國家文化記憶庫中尋找到所需要的資料。為解決這個問題，解題團隊的解題構想聚焦於利用語音輸入及自然語言文字機器人來協助無文化類專業知識背景的一般民眾，在搜尋國家文化記憶庫網站資料時，不會因為對專業用語的不了解，而找不到他們有興趣的資料。  解題團隊設計的系統允許使用者使用語音輸入法或是鍵盤輸入之自然語言查詢問句，透過NLP技術找出使用者意圖，並給予最相關的結果。在一般情形下，當使用者輸入查詢問句後，**GT搜尋法**可以依使用者輸入的問句主題自動衍生出更多其他使用者也會合併使用之關鍵字來精準找到符合使用者意圖的文章。如果使用者因不熟悉文化用語關鍵字，以致找不到符合使用者意圖的文章時 ，解題團隊提供了**FW搜尋法**，透過模糊比對技術，直接比對完整問句與國家文化記憶庫內文章主題關鍵字，即使輸入關鍵字不精確，也能找到近似結果。例如: 當使用者看完電影【賽德克巴萊】之後，透過國家文化記憶庫網站尋找【賽德克巴萊】，會找不到或只有很少的資料。但資料庫內實際上有【賽德克人】，【賽德克族】及【賽德克語】等豐富資料。這些相關資料便可以被FW搜尋法找到，即使使用的關鍵字不精確。最後，當使用者對文章的某一個句子有興趣，想要查詢更多相關資訊時，可以不用想如何輸入查詢問句。直接將整個句子複製到查詢區域，使用**ST搜尋法**即可利用句向量技術直接進行句子對句子的語義搜尋，找出語意相近之相關文章。  因此，整合解題團隊所開發的三種搜尋法到國家文化記憶網站中可以幫助國家文化記憶網站，除了提供很棒的使用者介面(UI)之外，貼心的搜尋機制，也能讓使用者有更好的使用者體驗(UX)。  二、解題技術架構及執行方法簡介  配合不同的使用者查詢場景，解題團隊提出三種搜尋法，分述如下:  (一) GT搜尋法 (Google Trend搜尋法)   1. GT搜尋法將使用者查詢問句經過CKIP斷詞及POS詞性分析後，進行NER實體名稱標註找出問句中最重要的NER關鍵字代表使用者意圖。再將NER關鍵字送入Google Trend裡面進行相關主題搜尋，取得熱搜衍生關鍵字集，最後再NER關鍵字、CKIP POS分析辭性為名詞的關鍵字以及Google Trend衍生關鍵字整合必對文章，並計算文章分數，得出排名。   (二) FW搜尋法 (FuzzyWuzzy搜尋法)  1. 預先對國家文化記憶庫公開資料集中所有的文章，找出文章NER關鍵字用來代表文章主題。  2. 將所有文章的NER關鍵字集合起來成為NER關鍵字集合。(解題團隊處理資料的過程中發現，類似的概念可以會因為描述的方式不同，而出現不同的NER關鍵詞，例如描述賽德克族的文化資料中可能出現賽德克、賽德克人等名詞，這些不同的描述方式都會讓使用單純字串比對的搜尋方式無法找到全部的資料。)  3. 使用FuzzyWuzzy比對使用者問句及文章NER關鍵字集，可以找出與問句相近數個NER關鍵詞及其相似度分數。  4. 以相近NER關鍵詞對所有文章進行字串比對，並相似度分數當成權重，即可算出每篇文章與問句的相符分數。依分數排名及可找出與使用者意圖相近之文章(即使使用者問句的關鍵詞不精確)  (三) ST搜尋法 (Sentence Transformer)  1. 預先將國家文化記憶庫公開資料集中所有的文章進行斷句處理。斷句的處理方式可以常見的標點符號如:。, ！, ？等符號來分割文章成多個句子。  2. 透過Sentence Transformer技術將文章中的每個句子都轉成對應的句向量。(不論句子長短，句向量的長度都相同，128個浮點數。句向量可視為句子中所有詞的詞向量經轉換後得出。因語相近的句子內含語意相近的詞，因詞向量的距離較晚，句向量的相似度也比較高)。  3. 當使用者輸入問句後，將問句也轉成句向量。透過比對問句向量與文章內句子之句向量相似度，並依相似度分數排名，即可得到語意最接近之句子及其所屬的文章。  三、成果與完成之工作目標  (一) 完成國家文化記憶庫公開資料集資料爬取及整理。  (二) 完成句子斷詞、詞性分析及NER關鍵字技術開發。  (三) 完成模糊比對技術開發與測試。  (四) 完成Google Trend API功能測試。  (五) 完成句向量模型設計、訓練與效能評估。  (六) 完成GT搜尋法設計、實作、測試與效能評估。  (七) 完成FW搜尋法設計、實作、測試與效能評估。  (八) 完成ST搜尋法設計、實作、測試與效能評估。  (九) 完成前端介面與後端程式的設計、實作、整合與測試。    圖1: 系統執行畫面  四、商轉可行性與成果創新價值  幾乎所有的公司都有自己的網站，也設置有搜尋引擎允許使用者以關鍵字查找內部網站資訊。解題團隊開發出的三種自然語言資料搜尋法可用來優化搜尋引擎，幫助使用者(1)透過GT搜尋法找到更精準的內容，(2)透過FW比對法找到近似的內容, (3)透過ST搜尋法以句子對句子語意搜尋的方式找到更多語意相近的內容。   1. 商轉可行性 2. 適用範圍廣 3. 本方案可提供政府或民間企業網站基於語意之自然語言形態內容搜尋服務，適用範圍廣泛。 4. ST搜尋法可協助政府及民間企業管理內部語意相似之文件。 5. 市場需求高   多數網站提供搜尋功能尋找網站資料，可提供之搜尋效能，市場需求高。   1. 系統具高移植性 2. 系統採用python為主要語言，移植性高。 3. 本系統可包裝成RESTful API與企業介接使用，不影響內部系統運作。 4. 系統擴充性高，可動態增加文章內容。 5. 商業模式   本方案開發之搜尋引擎效能提升方案可技術轉移給新創公司或資訊整合公司，被服務對象可一次支付一定的金額買斷技術或支付少量的金額可以在一定時間段內完全免費使用，期滿按周或月收費。   1. 技術創新亮點說明 2. 傳統的搜尋引擎透過字串比對或子字串比對找尋文章內容，使用者所使用之關鍵字品質影響搜尋結果。因此，無經驗的使用者常找不到需要的資料。為解決上述困難，解題團隊提出GT搜尋法、FW搜尋法及ST搜尋法。 3. GT搜尋法以問句NER關鍵字為主題，結合Google Trend找出其他使用者對NER關鍵字進行組合搜尋時所採用之文字，透過更多相關關鍵字之組合可精確找出符合使用者意圖之文章。 4. FW搜尋法將使用者輸入之完整問句與網站文章之NER主題關鍵字進行模糊比對，可找出與使用者意圖相同或最接近之相似NER主題關鍵字集合。透過相似NER主題關鍵字集合搜尋文章可找出與使用者意圖相同或相近之文章及主題。 5. ST搜尋法將網站文章句子轉成句向量，語意相近之句子會有距離相近之句向量。 ST搜尋法將使用者問句轉成句向量，透過比對使用者問句句向量與網站文章句向量之相似度，來提供基於語意之文章搜尋。   五、結論與建議  本方案對於內容網站所提出之GT搜尋法、ST搜尋法及FW搜尋法，具有相當高的實用性，且具有易建置、易維護及易移植之特性，所採用之方法具有理論的支持。解決了文化部國家文化記憶庫使用者搜尋找不到合適內容之問題，解題團隊提出之三種搜尋法可有效緩解使用者搜尋問題。提出之GT搜尋法可依照使用者意圖找到精確之文章內容。提出之FW搜尋法可在不確定關鍵字正確性之情形下找出網站內具有相近關鍵字之文章。提出之ST搜尋法可協助使用者以文章內之句子進行語意搜尋，幫助使用者衍生搜尋出更多感興趣之內容。建議之處為簽約時間、實作時間和報告時間稍嫌太短，希望能將時間給予更長時間以及提供更多的雲端運算資源，使訓練時間能夠加快速度。 | | | | |
| 解題成果  簡述 | | 量化  成果 | 1. GT搜尋法可利用短問句來進行查詢，不是每個民眾都使用專業術語進行查詢，當民眾利用短問句來進行查詢時，GT搜尋法會自動增加多數使用者會使用的相關關鍵字協助查詢，使得找出的文章更符合使用者意圖。因目前尚沒有客觀的方法可以衡量搜尋結果是否符合使用者意圖。因此，本系統採用主觀的方式評量搜尋結果是否符合使用者意圖。GT搜尋法找到文章的前十答案依使用者主觀判定符合意圖的比例為82.25%，前五排名依使用者主觀判定符合意圖的比例為是90%。 2. FW搜尋法可以將民眾輸入錯誤的文字時，進行修改並且回答出正確的文字，而非民眾原本輸入錯誤的文字，且列出與主題相關的文章。本系統所查詢的前十答案依使用者主觀判定符合意圖的比例92.8%，前五排名結果依使用者主觀判定符合意圖的比例為94.8%。 3. 本系統所提供的最後一套ST(句向量)查詢系統，其可以利用輸入較長的文章句子來進行查詢，本系統亦可以配合第一套Google Trend系統進行連鎖查詢，首先將利用第一套系統進行查詢，若沒有特別準確的答案，亦可以將查詢出來的句子放入句向量查詢，進而得到更為精準的答案。本系統產生的前十排名句向量答案依使用者主觀判定符合意圖的比例為59%，前五排名句向量答案依使用者主觀判定符合意圖的比例為88%。 | | | |
| 質化  成果 | 1. 本團隊所研發的GT搜尋法於第一次與出題單位線上會議展示時，初步搜尋結果符合出題單位期待，找到的文章內容較現有國家文化記憶網站的搜尋結果更符合使用者意圖。系統已安裝於雲端環境，出題單位可以隨時進行測試。 2. GT搜尋法的優點是: 即使使用者查詢文句中的關鍵字只有一個，但藉由Google Trend珍對該關鍵字所提供的相關主題可以得知現階段多數Google 使用者在查詢這個關鍵字時還會連帶查詢哪些關鍵字。這個線索可以幫助系統取得當下該關鍵字在多數人心目中的相關意圖可能有那些。藉由這些熱門的衍生關鍵詞協同搜尋，系統回應的文章有更高的機率找出符合使用者意圖的文章。 3. FW搜尋法的優點有二。第一，不對查詢問句進行斷詞，因此因斷詞錯誤而找不到答案的情形不會發生。第二，透過模糊比對技術可以將網站內所有NER主題關鍵字與問句相近之文章可以被找出來，即使使用者的關鍵字不完全與NER關鍵字相同。 4. ST搜尋法的優點是: 非常適合以文章內文查詢其他內文。ST搜尋法適合較長句子的查詢問句，但一般使用者輸入的較長查詢問句卻不一定回找到符合意圖的答案。反而是透過文章內文查詢其他文章內文會有比較高的機率找到符合使用者意圖的文章。經判定可能的原因是，一般使用者查詢問句的用詞與語法結構與文字工作者的用詞及語法還是有很大的差距，這也是為何國家文化記憶庫的使用者如不具備專業知識，經常會找不到符合意圖資料的可能原因。 | | | |
| 成果電子檔 | | <https://203.145.219.146:59757/>  https://youtu.be/HN3QLrJAyjI | | | | |
| 備註 | |  | | | | |

1. 題目背景說明

**(一)、**題目理解說明:

一般民眾對於文化類的資料及專業術語較為陌生，難以利用精確的關鍵字透過文化部網站中找到符合意圖的資料。 文化部國家文化記憶庫網站的資料是由專家整理或者是專家所提供的，且素材內容有特定的主題，但普羅並不一定具備這些主題的相關知識，也無法以正確的主題關鍵字找到所需的素材。此外，民眾若使用自然語言的方式來查詢，也可能會因為查詢語句中沒有具體的關鍵字，而無法找到特定主題的素材。因此，如果系統能允許使用者輸入意義相近的關鍵字或是意義相近之語句，並自動協助使用者找到相近專業文化類關鍵詞，並能找到相對應之素材，使用者會更加願意去使用系統來查詢他們不了解的素材。

解題構想聚焦於利用語音輸入及自然語言文字機器人來協助無文化類專業知識背景的一般民眾，能在透過國家文化記憶庫網站查詢資料時，不會因為專業背景不足，而找不到符合使用者意圖的資料。透過自然語言處理技術的應用，解題團隊期待使用者有很大的機會可以找到符合使用者查詢意圖的資料，即使只用了一個關鍵字，甚至是使用一個不精確的關鍵字。解題團隊也期待當使用者找到感興趣的文章之後，可以進一步利用該文章的內文，以句子找句子的方式，找出國家文化記憶庫中，語意內容與本文章相近的文章也能一併被找出來。

(二)、構想背景說明：

當使用開始搜尋資料時，其心中必有意圖想要找尋某一類資料。假設使用者能以特定的關鍵詞來描述其意圖，此一意圖應該是特定的人、事、時、地、物。當使用者使用自然語言描述其意圖時，其關鍵字很有會以命名實體名稱(Named Entity Recognition, NER)的形式出現在查詢語句中。因此，找到NER即可找到使用者意圖。但因為NER關鍵字出現在一篇文章中，不一定代表該文章的主題與NER關鍵字相對應。但如果能找到與NER關鍵字相關之衍生關鍵字，並用來找出同時包含NER關鍵字與衍生關鍵字的文章，則找到的文章與使用者意圖相符的文章機會將大為增加。基於此一觀察，解題團隊提出【GT搜尋法】可以取出使用者查詢文句中的使用意圖，並將該意圖透過Google Trend找出更多衍生關鍵字。將使用者意圖關鍵字(NER關鍵字)、查詢問句中出現的名詞及Google Trend衍生出之關鍵字綜合比對所有文章進行比對分數計算，並將比對分數由高到低排序，即可得到符合使用者意圖之文章排序。

如果使用者因不熟悉文化用語關鍵字，以致找不到符合使用者意圖的文章時 ，解題團隊提供了【**FW搜尋法】**，透過模糊比對技術，直接比對完整問句與國家文化記憶庫內文章主題關鍵字，即使輸入關鍵字不精確，也能找到近似結果。例如: 當使用者看完電影【賽德克巴萊】之後，透過國家文化記憶庫網站尋找【賽德克巴萊】，會找不到或只有很少的資料。但資料庫內實際上有【賽德克人】，【賽德克族】及【賽德克語】等豐富資料。這些相關資料便可以被FW搜尋法找到，即使使用的關鍵字不精確。

最後，當使用者對文章的某一個句子有興趣，想要查詢更多相關資訊時，可以不用想如何輸入查詢問句。直接將整個句子複製到查詢區域，使用【**ST搜尋法】**即可利用句向量技術直接進行句子對句子的語義搜尋，找出語意相近之相關文章。

因此，整合解題團隊所開發的三種搜尋法到國家文化記憶網站中可以幫助國家文化記憶網站，除了提供很棒的使用者介面(UI)之外，貼心的搜尋機制，也能讓使用者有更好的使用者體驗(UX)。

(三)、解決方案之文獻、市調資料回顧(以補齊)

1.本團隊的GT搜尋系統是參照 “Discovering Trends in Software Engineering with Google Trend”這篇論文的想法進行解決方案之執行，論文網址如下: <https://dl.acm.org/doi/epdf/10.1145/1234741.1234765>。

在2006年五月，Google Trend已經開放給大眾進行使用所想要查詢的主題，其功能有分析媒體所關注的議題、找尋興趣及科技、工具及方法的因果關係。因此本團隊發現，竟可以利用一個關鍵字來衍生查詢有關於其他跟關鍵字有關聯的主題。所以將此方法用於將文化部內所有文章進行關鍵字的標註，後利用Google Trend的熱門搜尋字來衍生另一類關鍵字，使用者進行查詢時，可以將所有相關的主題給予使用者，進一步解決不精確的主題問題。

2. 本解題方案的FuzzyWuzzy, FW搜尋法參考文獻為 IEEE Conference Publication | IEEE Xplore”Dynamic Fuzzy String-Matching Model for Information Retrieval Based on Incongruous User Queries” 此文獻運用推理策略使訊息檢索能減少 不匹配的可能性，並提出了一個模糊字符串匹配模型來協助使用者時常因查詢不一致可能導致不確定性、錯誤信息、誤傳、缺乏足夠的領域知識等，此模型也幫助使用者根據模糊推理模型策略於查詢文件效果更好。

3. 本解題方案的句向量，ST搜尋系統的參考文獻為” Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks” 這篇論文之想法進行解決方案之執行，論文網站如下:  [**Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks | arXiv:1908.10084**](https://arxiv.org/abs/1908.10084)

**表單的頂端**BERT在語義文本相似性 (STS) 等句子對回歸任務上具有相當先進的性能，在處理兩個句子的相似度計算也包含在其中，利用其相似度計算出分數再進行比較。過往，在查詢系統上沒有使用句向量的方式來進行查詢，因此本團隊注重利用句向量來進行處理，其查詢速度及精準度皆有良好的表現，代表此查詢系統方法具有一定的未來性及發展性，利用文章內文查詢相關資訊，並且衍生有關之主題，提供使用者不同面向的參考。

**二、解題技術架構及執行方法**

本團隊的解題構想，將利用自然語言及深度學習的技術，將文化部國家文化記憶庫網站上面的資料視為輸入資料，透過三種不同的搜尋法來進行本專案的查詢系統。

GT搜尋法(符合使用者意圖)是將使用者查詢問句經過斷詞斷句以及POS詞性分析後，進行NER實體標註關鍵字，取得關鍵字後，送入Google Trend裡面進行熱門關鍵字搜尋，取得熱搜衍生關鍵字集，最後再利用字串比對來計算分數，得到排名。

FW搜尋法首先將所有文章的NER關鍵字處理好並集合成關鍵字集，當使用者輸入之完整查詢問句，透過模糊比對技術比對完整查詢問句與文章關鍵字集比對之相似度得出模糊關鍵字集及模糊比對分數。使用模糊關鍵字集當查詢關鍵字並以模糊比對分數當成權重，對文章進行比對與分數計算。再依照分數排序得出相關文章列表。

ST搜尋法預先將文化部國家文化記憶庫網站的所有文章進行標點符號、特殊符號的處理及切割，再將每句文章斷句送入Bert模型進行向量編碼。當使用者輸入查詢問句後，也將問句透過BERT模型轉成句向量。在句向量空間比對查詢問句與文章文句之相似度，相似度越高代表語意越接近。再將該文句所屬的文章取出回饋給使用者。

整體系統架構圖如圖2所示，前端為網頁介面允許使用者以語音或文字的方式輸入查詢問句並選擇搜尋的方式。系統將查詢問句送入後段網頁伺服器，並依使用者選擇的方式選用對應的搜尋法進行資料檢索，找到的資料再傳給網頁伺服器將結果以網頁的方式回傳給使用者。

|  |
| --- |
|  |
| 圖2: 系統架構圖 |

1. GT搜尋法典藏資料搜尋

GT搜尋法為系統預設的搜尋演算法。當使用者輸入查詢文句時，其意圖可以透過自然語言處理技術中的實體名稱辨識(Named Entity Recognition, NER) 技術來找到，並使用NER關鍵詞找到相關主題之文章。但因為文章內容的多樣性，文章內容出現NER關鍵字並不代表該篇文章的主題與使用者意圖相符。也有可能只是附帶提到的名詞。為了解決這個問題，GT演算法將NER關鍵詞帶入Google Trend搜尋更多的相關主題，當成衍生關鍵詞。因Google Trend的相關主題代表的含意是多數使用者在查詢NER關鍵詞時，會結合使用進行組合查詢的關鍵字。因此，查詢結果相關主題和帶入查詢的NER關鍵詞有密切的關連。可將相關主題視為NER關鍵字之衍生關鍵字。若A文章只出現NER關鍵詞，但B文章卻有NER關鍵詞及衍生關鍵詞。則B文章的內容比A文章內容有更大的機會符合使用者意圖。GT搜尋法處理流程如圖3所示。詳細處理步驟說明如下：

|  |
| --- |
|  |
| 圖3: GT搜尋法處理流程 |

1. 使用者輸入問句後，系統將會使用CKIP對問句進行斷詞以及Part of Speech (POS)詞性分析，再利用斷詞結果及POS詞性分析找出NER關鍵字。NER關鍵字及POS詞性為名詞之斷詞皆為搜尋符合使用者意圖文章之重要關鍵詞。
2. 將取出之NER關鍵字輸入Google Trend進行查詢，可取得該關鍵字的熱門搜索相關主題，可將相關主題關鍵字集合成NER衍生關鍵字。如圖4所示。
3. 最後將NER關鍵字及CKIP名詞斷詞，及NER衍生關鍵字對文章進行不同權重組合比對即可計算出一篇文章符合使用者意圖的分數。
4. 再將文章依分數由高至低排序及可得到符合使用者意圖之文章排名。

|  |
| --- |
|  |
| 圖4: 將問句中之NER關鍵字透過Google Trend產生多數搜尋者會使用之關鍵字組合，以找出更精確的文章內容。 |

圖5為GT搜尋法搜尋範例。當使用者輸入查詢問句【我要查詢孫中山先生】，透過CKIP技術對問句進行斷詞，可以得到NER關鍵字【孫中山】及CKIP名詞【先生】。再將NER關鍵字【孫中山】帶入Google Trend查詢可得到相關主題【國父】【後代】【辛亥】【中國】【三民主義】…等詞。將NER關鍵詞權重設成100，CKIP名詞權重設成10，NER衍生關鍵字權重設成1，對文章進行符合使用者意圖分數計算，得到某篇文章符合使用者意圖分數為224分。將文章依分數由高至低排序即可將最符合使用者意圖的文章列表回應使用者。

|  |
| --- |
|  |
| 圖5: GT搜尋法搜尋範例 |

1. FW搜尋法典藏搜尋

FW搜尋法的主要目的是透過模糊比對技術協助使用者盡量找到文章NER關鍵字集中與查詢問句最近似之模糊關鍵詞及模糊比對分數。再使用模糊關鍵字集並使用模糊比對分數當成權重，對文章進行符合使用者程度計分。越精準的關鍵字權重越高，越不精準的關鍵字權重越低。最後再依照分數由高至低排序即可得到最符合使用者意圖之文章。FW搜尋法的長處是即使查詢問句中的關鍵詞與文章關鍵詞不完全一樣，文章還是有可機會被找出來。FW搜尋法執行流程如圖6所示。FW搜尋法執行流程說明如下:

1. 使用爬蟲程式蒐集文化部國家文化記憶庫網站公開資料集內所有的文章資料。並以資料集中，文章裡的Description欄位內容進行斷詞處理。

|  |
| --- |
|  |
| 圖6: FW搜尋法處理流程 |

1. 接著繼續利用CKIP進行NER處理，取出之NER類別可分成3大類(實體、時間、數字)7小類(人名、地名、組織機構名、時間日期、貨幣、百分比)，考量使用者可能的問句查詢意圖，FW搜尋法只取出人名、地名、組織機構名以及時間日期，並將結果輸出成Json檔案儲存。如圖7所示。

|  |
| --- |
|  |
| 圖7: 將文化記憶庫文章所提取的NER關鍵字整理成關鍵字集供後續模糊比對之用。 |

1. 當使用者輸入查詢問句時，將完整問句與NER關鍵字集進行Fuzzy Wuzzy比對後，找出與問句最相關之模糊NER關鍵字集及相似度分數。
2. 最後使用模糊關鍵字集及相似對分數對所有文章進行組合比對即可算出每篇文章符合使用者意圖的分數。將分數由高到低排序即可找出最符合使用者意圖之文章。

|  |
| --- |
|  |
| 圖8: FW搜尋法搜尋範例  圖8為FW搜尋法搜尋範例。假設使用者看完電影後想要搜尋更多有關電影【賽德克巴來】相關資料，當使用者輸入【我想要查賽德克巴來】之後，FW搜尋法將查詢問句與資料庫中之NER關鍵字集比對，找到【賽德克】，【賽德克人】，【賽德克族】及【賽德克語】等模糊NER關鍵詞，及其模糊比對分數。再使用模糊NER關鍵詞，並以模糊比對分數當成權重，對所有文章計算符合使用者意圖分數。最後再將符合使用者意圖分數由高到低排序即可得到最符合使用者意圖之文章列表。 |

1. ST搜尋法

在自然語言深度學習模型中，Sentence Transformer被訓練成只要輸入之兩個句子語意相近，這兩個句子轉換出之句向量便會有比較高的相似度。基於此一觀察，ST搜尋法使用。, ！, ？等標點符號將國家文化記憶庫中所有的文章都分割成多個句子，並將每個句子都轉換成句向量並記錄文章與句向量之關係。當使用者輸入查詢問句後，也將查詢問句透過Sentence Transformer轉換成句向量。透過比對查詢句向量與文章句向量之相似度，並計算出相似度分數。只要將所有的句子依照句向量相似度由高到低排序，即可找出與查詢問句最接近之句子及其所屬之文章。ST搜尋法分成訓練期及使用期，其運作方式如圖10所示。ST搜尋法之詳細處理步驟說明如下:

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 圖10: ST搜尋法之訓練期及使用其流程 | |
|  |

1. 通過爬蟲程式收集國家文化記憶庫網站內所有的公開資料集文章。並利用語句結尾標點符號對文章進行處理及切割，形成文章斷句。
2. 將文章斷句送入BERT模型將斷句轉換成長度為128之句向量編碼，並將處理好的向量編碼儲存成json檔案。
3. 當使用者輸入查詢問句時，ST搜尋法會透過BERT模型算出問句句向量編碼。將問句句向量與預先計算好之文章句向量編碼進行餘弦相似度比對。語意相近之句子配對會有比較高之餘弦相似度值。將相似度排序由高至低排序。高相似度句向量之原始句子及其所屬文章即為符合使用者意圖之文章。比對過程如圖12所示。

|  |
| --- |
|  |
| 圖11: 將查詢文句轉成句量，透過比對查詢問句句向量即文章文句句向量之相似度即可找到包含相近語意文字之文章。 |

**三、成果效益與完成之工作**

1. 完成之工作目標
2. 完成國家文化記憶庫公開資料集資料爬取及整理。
3. 完成句子斷詞、詞性分析及NER關鍵字技術開發。
4. 完成模糊比對技術開發與測試。
5. 完成Google Trend API功能測試。
6. 完成句向量模型設計、訓練與效能評估。
7. 完成GT搜尋法設計、實作、測試與效能評估。
8. 完成FW搜尋法設計、實作、測試與效能評估。
9. 完成ST搜尋法設計、實作、測試與效能評估。
10. 完成前端介面與後端程式的設計、實作、整合與測試。

|  |
| --- |
|  |
| 圖11: 語意機器人典藏搜尋系統 |

1. 量化成果

考量目前並沒有客觀的方法可以評量查詢結果是否符合使用者意圖。為評量系統效能，解題團隊採用使用者主觀評斷查詢結果是否符合其搜尋意圖的方式來評量系統效能。

1. **GT搜尋法效能評估:**

GT搜尋法為系統預設之搜尋法，使用者可利用短問句來進行查詢。因GT搜尋法可以尋找更多相關的衍生關鍵字來偕同搜尋，即使使用者的問句是很簡單的單一關鍵字，搜尋結果前幾名還是比傳統的字串比對結果有更高的機會滿足使用者主觀判定符合使用者意圖。解題團隊使用GT搜尋法讓測試者輸入20個問題，並將搜尋結果的前10名及前5名結果依使用者主觀判定是否符合使用者意圖的方式計算符合率。GT查詢法得到答案的的前10名結果主觀判定符合使用者意圖的比率是82.25%，前5名結果主觀判定符合使用者意圖之比率為90%。表1及表2為兩個GT搜尋法搜尋範例，表上方為使用者查詢問句，關鍵字欄位內容為該篇文章內出現符合問句NER關鍵字、CKIP名詞關鍵字或使用Google Trend找出之NER衍生關鍵字之文章關鍵字。

表1: GT搜尋法搜尋範例1

|  |  |  |
| --- | --- | --- |
| 問句:媽祖 | | |
| 關鍵字 | 文章主題 | 主觀判定是否符合使用者意圖。 |
| 進香，神，媽祖 | 天祝宮\* | 是 |
| 神，媽祖 | 媽祖巨神像，世界之最 | 是 |
| 神，媽祖 | 進安宮\*\* | 是 |
| 馬祖列島，神，媽祖 | 南竿馬祖境天后宮 | 是 |
| 神，媽祖 | 王天君 | 是 |
| \*天祝宮:台灣唯一一座站立媽祖神像的廟宇。  \*\*進安宮:進安宮內開基媽祖擁有兩百多年歷史。 | | |

表2: GT搜尋法搜尋範例2

|  |  |  |
| --- | --- | --- |
| 問句:日治時期 | | |
| 關鍵字 | 文章主題 | 主觀判定是否符合使用者意圖。 |
| 臺灣，日本，政府，建築，日治時期，時期 | 日治時期基隆「憲兵隊」 | 是 |
| 政府，日治時期，時期 | 社寮島、八尺門、大沙灣略圖\* | 是 |
| 臺灣日治時期，臺灣，日本，文化，建築，日治時期，時期 | 豐田移民村醫生住家舊宅 | 是 |
| 日本，戰爭，日本，日治時期，歷史 | 日治時期─南機場(飛機跑道) | 是 |
| \*大沙灣略圖:日治時期的地名標註。 | | |

1. **FW搜尋法效能評估:**

FW搜尋法在使用者輸入查詢問句後，會保留完整問句，不經斷詞處理，並使用FuzzyWuzzy方法比對文章NER關鍵字集與查詢問句的相似度。評量效能的方式也是使用使用者主觀判定符合使用者意圖的方式進行。使用者輸入20個不同的問句，取出FW搜尋法找出的前10名及前5名結果進行評估。每個問句的前10名結果使用主觀判定符合使用者意圖的比率是92.8%，每個問句的前5名結果使用主觀判定符合使用者意圖的比率是94.8%。 表3及表4呈現的是FW搜尋法搜尋範例。表格上方為查詢問句及比對文章NER關鍵字及查詢問句後得到分數較高之模糊關鍵字及比對分數。文章關鍵字欄位呈現的是該篇文章中出現之模糊關鍵字。

表3: FW搜尋法搜尋範例

|  |  |  |
| --- | --- | --- |
| 問句:宋元明清歷史故事  模糊關鍵字: ('元明', 90), ('宋元', 90), ('明清', 90), ('故事館', 72), ('宋哲元', 60) | | |
| 文章關鍵字 | 文章內對應文句 | 主觀判定是否符合使用者意圖。 |
| 明清，宋元，元明 | 周澄曾侍近現代書畫名家江兆申（1925-1996）研習書畫篆刻兼讀史書詩詞，繪畫創作取涉宋元明清歷代山水大家，包括清代石濤（1642-1718）等名家，博採眾長，融古出新。 | 是 |
| 明清 | 討論眷村改建保留與活化再生的「眷村文史座談」、明清及日據時代的臺灣眷村講座；於電視節目播出的「眷村訪查報導」及「表演藝術季」的劇團演出。 | 是 |
| 明清 | 芹壁聚落的開發，最早可追溯至明清時期，此地多作為中國大陸沿海漁民海上停泊休憩之地。 | 是 |
| 宋元 | 我鄉各島羅列東海，宋元以降不乏八閩丁壯渡海棲，止櫛沐胼胝或漁或耕夙夕慘淡於斯，營生先民敬天法祖尊神崇祇，其對天后神靈尤為頂禮，帡幪張覆無論除災去診抑或航海梯山無不四境蔭佑使民大和償安。 | 是 |

表4: FW搜尋法搜尋範例2

|  |  |  |
| --- | --- | --- |
| 問句：皇民化運動時期相關人物  模糊關鍵字:('皇民化', 90), ('皇民化運動', 90), ('昭和12年（1937）4月皇民化運動時期', 70), ('昭和12年（1937）皇民化運動時期', 70), ('皇民化時期', 62) | | |
| 文章關鍵字 | 文章內對應文句 | 主觀判定是否符合使用者意圖。 |
| 皇民化, 皇民化運動, 昭和12年（1937）4月皇民化運動時期 | 拔子公學校創立於大正2年（1913）4月1日，坐落於漢人聚落拔仔庄內。拔仔庄在清帝國後期便有軍隊進駐，並有移民隨之前來開墾，為花東移民史上最早開發的區域之一。昭和12年（1937）4月皇民化運動時期，拔子庄改為白川，校名也隨地名更為「白川公學校」。 | 是 |
| 皇民化, 皇民化運動, 昭和12年（1937）皇民化運動時期 | 大庄為今日花蓮縣富里鄉東里村的舊稱。此地居民多是西拉雅族人從南臺灣東遷至舊庄（今玉里長良），後續再往秀姑巒溪東岸遷至今日東里村一帶形成大部落，而稱為大庄。昭和12年（1937）皇民化運動時期將大庄改名為大里。 | 是 |
| 皇民化, 皇民化運動, 昭和12年（1937）皇民化運動時期 | 昭和12年（1937）皇民化運動時期，日本人把織羅、馬太林兩部落合併改稱「春日」，沿用至今。織羅派出所與觀音山派出所一樣都位於秀姑巒溪東側的海岸山脈山麓地帶，主要工作是原住民的「教化」，另外也擔任青年團、壯丁團、夜學會的指導，當時有壯丁團員50名、青年團員65名以及夜學會員87名。 | 是 |

1. **ST搜尋法效能評估:**

ST搜尋法使用句向量的方式比對查詢問句與文章內句子句向量相似度。ST搜尋法適合輸入較長的文章句子來進行查詢。經實驗測試，使用文章內文句子，以句子查句子的方式查詢會比使用者輸入的問句有更好的效果。評估可能的原因是文字工作者的用字與筆觸與一般使用者還是存在著差異。建議可以先使用GT搜尋法找出感興趣的內文，再使用ST搜尋法以句子查句子的方式查詢會有比較好結果。使用者輸入20個問句進行測試，所有的測試樣本都是由每一篇文章的內文以複製貼上的方式進行搜尋。經主觀判定是否符合使用者意圖，前10名答案依主觀判定符合使用者意圖的比率只有59%，但如果範圍縮小前5名的答案，則主觀判定符合使用者需求的比例可以大幅提升到88%。表5及表6為ST搜尋法之搜尋範例，表上方為使用者查詢問句，相似度欄位為查詢問句句向量與相符文句句相樣之餘弦相似度，相符文具則是文章內文中與查詢問句語意最相近的句子。

表5: ST搜尋法搜尋範例1

|  |  |  |
| --- | --- | --- |
| 問句: 兒童文學創作，內容簡單易懂，以小孩子第一次就讀幼稚園的經驗為主題 | | |
| 相似度 | 相符文句 | 主觀判定是否符合使用者意圖。 |
| 0.859093 | 此書為方梓第二本兒童文學創作，內容簡單易懂，以小孩子第一次就讀幼稚園的經驗為主題，書寫小孩面對嶄新環境的心境，也試著鼓勵年幼的讀者群，不必懼怕接觸新環境。 | 是 |
| 0.714937 | 其中最廣為人知的是兒童文學的創作，曾連續獲得洪建全兒童文學創作獎少年小說首獎。 | 是 |
| 0.707403 | 對於兒童們學習美語，培養閱讀興趣，促進親子共讀，這裡猶如一座寶山。 | 是 |

表6: ST搜尋法搜尋範例2

|  |  |  |
| --- | --- | --- |
| 問句: 棒球運動何時傳入台灣？ | | |
| 相似度 | 相符文句 | 主觀判定是否符合使用者意圖。 |
| 0.8591792 | 棒球運動何時傳入東台灣？根據湯川充雄《臺灣野球史》一書中記載，大正6年（1917）台東就已經有了製糖團等企業團體或政府單位所組成的棒球隊，但是根據圖中這張林桂興所遺留下的珍貴歷史影像顯示：在大正4年（1915）時，台東就已經有正規棒球隊的存在了 | 是 |
| 0.800838 | 《台灣野球史》一書中曾經記錄昭和6年（1931）11月宜蘭、花蓮港曾經進行棒球友誼賽，但是比賽地點在花岡山，所以並不是這張照片所記錄的這一場 | 是 |
| 0.766628 | 棒球在日治時期傳入臺灣，臺灣第一支棒球隊──臺灣總督府中學校棒球隊──創立於1906年（明治39年）；戰後初期至1960年代花蓮市民對棒球的愛好與日俱增，1958年（民國47年）市長盃棒球賽在玉里隊與鐵路隊舉行冠亞軍決賽之前，大會應球迷要求特別安排一場胖子隊與瘦子隊的趣味競賽，胖瘦之分以體重超過120台斤（相當於72公斤）為判 | 是 |

1. 質化成果
2. 解題團隊與文化部進行第一次交流會議時，展示了GT搜尋法。文化部資訊處長官認為已比目前運行系統的效能好多了。解題團隊的解讀是搜尋結果依主觀判定符合使用者需求的比率必目前運行的系統更高。
3. GT搜尋法的優點是: 即使使用者輸入單一關鍵字，GT搜尋法仍會設法找到更多的相關關鍵字，並以組合查詢的方式找到更符合使用者意圖之文章。
4. FW搜尋法能將使用者輸入的問句以模糊比對的方式找出相近的內容。
5. ST搜尋法可以提供使用者以句子查詢更多語意相關句子的功能。使用者可以輸入長度較長的句子，句子越長，找出語意相近句子的機會增加。

**四、商業應用價值與創新亮點**

1. 解題方式商轉可行性
2. Google Trend搜尋系統可以使用於當使用者不明確自己所想要查詢的主題時，此系統可以將使用者輸入的文字將其衍生出多樣的主題性，使用者即可以慢慢選取所想要的主題。例如: 電商，當一位顧客想要尋找東西卻沒有任何想法，此系統可以將顧客輸入的任何文字進行衍生，進而查詢到多樣性的資料供顧客選取。
3. 模糊比對搜尋系統可以將使用者輸入不精確的文字進行修改並且將其有關係的文章列表出來，既可以不用擔心是否自己輸入的文字是否正確，此系統還是可以輸出與使用者想要的資料有關的訊息文章。例如: 各大查詢系統，使用者輸入錯誤的關鍵字或輸入文字有缺少文字時，此系統將可以立即的修正錯誤的文字，並且列出與輸入文字有關的主題及資料提供給使用者。
4. 句向量搜尋系統可以用於長句子或文章內文句子，使用者輸入字數較長的句子或輸入書本內的句子及網路文章的句子，其查詢速度之快以及精確度佳，可用於圖書館查詢系統，使用者可利用書本的書名或書本內的句子進行查詢，此系統將會輸出最符合書本相關的資料提供給使用者。
5. 另外，本系統也具備易建置、易移植及易維護等性能，在商業運轉上同時具有可行性。以下將條列說明本方案商轉的優勢:
6. 解決的痛點
7. GT搜尋法比原搜尋系統更貼近使用者搜尋意圖。
8. FW搜尋法比原搜尋系統更能處理不精準的文字查詢。
9. ST搜尋法提供原搜尋系統無法提供之語意衍生查詢服務。
10. 適用範圍廣
11. 本方案可提供政府或民間企業網站基於語意之自然語言形態內容搜尋服務，適用範圍廣泛。
12. ST搜尋法可協助政府及民間企業管理內部語意相似之文件。
13. 市場需求高

1. 多數網站提供搜尋功能尋找網站資料，可提供之搜尋效能，市場需求高。

1. 系統具高移植性
2. 系統採用python為主要語言，移植性高。
3. 本系統可包裝成RESTful API與企業介接使用，不影響內部系統運作。
4. 系統擴充性高，可動態增加文章內容。
5. 商業模式

1. 本方案開發之搜尋引擎效能提升方案可技術轉移給新創公司或資訊整合公司，被服務對象可一次支付一定的金額買斷技術或支付少量的金額可以在一定時間段內完全免費使用，期滿按周或月收費。

1. 技術創新亮點說明
2. 傳統的搜尋引擎透過字串比對或子字串比對找尋文章內容，使用者所使用之關鍵字品質影響搜尋結果。因此，無經驗的使用者常找不到需要的資料。為解決上述困難，馬訓冷凝團隊提出GT搜尋法、FW搜尋法及ST搜尋法。
3. GT搜尋法以問句NER關鍵字為主題，結合Google Trend找出其他使用者對NER關鍵字進行組合搜尋時所採用之文字，透過更多相關關鍵字之組合可精確找出符合使用者意圖之文章。
4. FW搜尋法將使用者輸入之完整問句與網站文章之NER主題關鍵字進行模糊比對，可找出與使用者意圖相同或最接近之相似NER主題關鍵字集合。透過相似NER主題關鍵字集合搜尋文章可找出與使用者意圖相同或相近之文章及主題。
5. ST搜尋法將網站文章句子轉成句向量，語意相近之句子會有距離相近之句向量。 ST搜尋法將使用者問句轉成句向量，透過比對使用者問句句向量與網站文章句向量之相似度，來提供基於語意之文章搜尋。

**五、結論**

本方案對於內容網站所提出之GT搜尋法、ST搜尋法及FW搜尋法，具有相當高的實用性，且具有易建置、易維護及易移植之特性，所採用之方法具有理論的支持。解決了文化部文化記憶庫使用者搜尋找不到合適內容之問題，解題團隊提出之三種搜尋法可有效緩解使用者搜尋問題。提出之GT搜尋法可依照使用者意圖找到精確之文章內容。提出之FW搜尋法可在不確定關鍵字正確性之情形下找出網站內具有相近關鍵字之文章。提出之ST搜尋法可協助使用者以文章內之句子進行語意搜尋，幫助使用者衍生搜尋出更多感興趣之內容。建議之處為簽約時間、實作時間和報告時間稍嫌太短，希望能將時間給予更長時間以及提供更多的雲端運算資源，使訓練時間能夠加快速度。

**六、附件(另含附件會議紀錄資料及照片集。)**

**附件1: 與出題單位會談之會議紀錄。**

1. 會議記錄

初次與企業討論的系統架構投影片

|  |
| --- |
|  |
| 圖12: 開會投影片 |

1. 內部專案討論及開會

|  |
| --- |
|  |
| 圖13: 內部開會情形 |

1. 會議記錄

|  |
| --- |
|  |
| 圖14: 第一次開會會議紀錄 |

|  |
| --- |
|  |
| 圖15: 內部開會會議紀錄 |
| 1. 與出題企業聯繫紀錄  |  | | --- | | 表7: 與出題企業聯繫紀錄 | |  | |

(三)解決方案創造之商業合作佐證(如：衍伸之商業合約、政府補助案遞件等證明)

**附件2: 與出題單位簽訂之保密協議**

|  |
| --- |
|  |
| 圖13: 與企業簽訂之保密協議 |

**附件3、獎補金核銷表**

自中華民國110年07月至110年10月

【請以解題獎補助金30萬元規模編列】

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 項目 | 內容細項 | 關聯性說明 | 單價  (含稅) | 數量 | 小計  (含稅) |
| 人事費 | 主管加給 |  | 40000 | 1 | 40000 |
| 薪資及午餐費 |  | 30000 | 4 | 120000 |
|  |  |  |  |  |
| 運算資源 | 電腦設備 |  | 5000 | 4 | 20000 |
|  |  |  |  |  |
|  |  |  |  |  |
| 差旅費 | 交通費 |  | 10000 | 4 | 40000 |
| 餐飲費 |  | 10000 | 4 | 40000 |
|  |  |  |  |  |
|  |  |  |  |  |
| 雜支 | 文具、影印、耗材、書籍採購 |  | 10000 | 4 | 40000 |
|  |  |  |  |  |
|  |  |  |  |  |
| 其他項目 |  |  |  |  |  |
|  |  |  |  |  |
| 總計(含稅) | | | | | 300000 |

參考文獻

**[1]** Jörg Rech “Discovering Trends in Software Engineering with Google Trend,” in [Volume 32](https://dl.acm.org/toc/sigsoft/2007/32/2),[Issue 2](https://dl.acm.org/toc/sigsoft/2007/32/2),March 2007 ,pp 1~2

**[2]** Olufade, F. W. Onifade, Member IAENG, Odile Thiéry, Adenike, O. Osofisan, Gérald Duffing,"Dynamic Fuzzy String-Matching Model for Information Retrieval Based on Incongruous User Queries", Proceedings of the World Congress on Engineering 2010 Vol I WCE 2010, June 30 - July 2, 2010, London, U.K.

**[3]** N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.