

中華民國第 62 屆中小學科學展覽會

作品說明書

高級中等學校組 行為與社會科學科

佳作

052702

由人工智慧生成吸引人的言情小說之可行性研究

學校名稱：新北市立林口高級中學

作者： 高二 陳俞辰	指導老師： 胡裕仁 潘尚怡
---------------	---------------------

關鍵詞：生成對抗網路、網路爬蟲、言情小說

摘要

近年來言情小說盛行，發展出各式各樣的模式。以台灣為例，目前言情小說的出版量，每年高達一千多本，佔全台小說出版總量之 3 至 4 成[5]。對出版社而言，如果可以了解為何言情小說能在眾多類型的書籍中吸引讀者，並找出言情小說成功的劇情寫作模式，將是一件很有商業價值的事情。人工智慧(Artificial Intelligence, AI)是目前產業當紅的科技技術，若能結合 AI 解決未來出版業的新書企劃，不但擁有商業價值，亦具有話題性題材。由於 AI 的效能好壞，取決於資料本身的品質及訓練資料的數量。但在實際應用中，資料取得往往費時費力且所費不貲。因此本研究採用人工分析及網路爬蟲雙軌模式進行，並使用生成對抗網路的 AI 模式，協助開發暢銷小說企劃情節，並找出暢銷劇情寫作核心。

壹、研究動機

自古至今言情小說一直在中、外各地盛行，例如：清代曹雪芹的紅樓夢；英國威廉·莎士比亞的羅密歐與朱麗葉等，演化至今已同時發展出各式各樣的文風與模式。同樣身為喜愛閱讀言情小說的我，期望能夠藉由分析相關資料，了解為何言情小說能在眾多類型的書籍中吸引眾多的讀者，並且結合現今的人工智慧，設計出可以供電腦自動生成暢銷的言情小說寫作劇情核心題材。

貳、研究目的、流程及架構

本研究計畫希望能夠藉由人工智慧方法的應用，找出言情小說成功的劇情寫作模式。茲將本計畫的研究目的整理如下：

- (一) 言情小說的主要寫作結構分析。
- (二) 吸引讀者的主要情節結構成因分析。
- (三) 歸納商業化言情小說情節及劇本設計模式。
- (四) 以對抗生成網路生成商業化言情小說情節及劇本設計模式之可行性評估。

依據上述之研究目的，整個研究計畫之研究流程與生成對抗網路研究架構，分別如圖 1 與圖 2 所示：

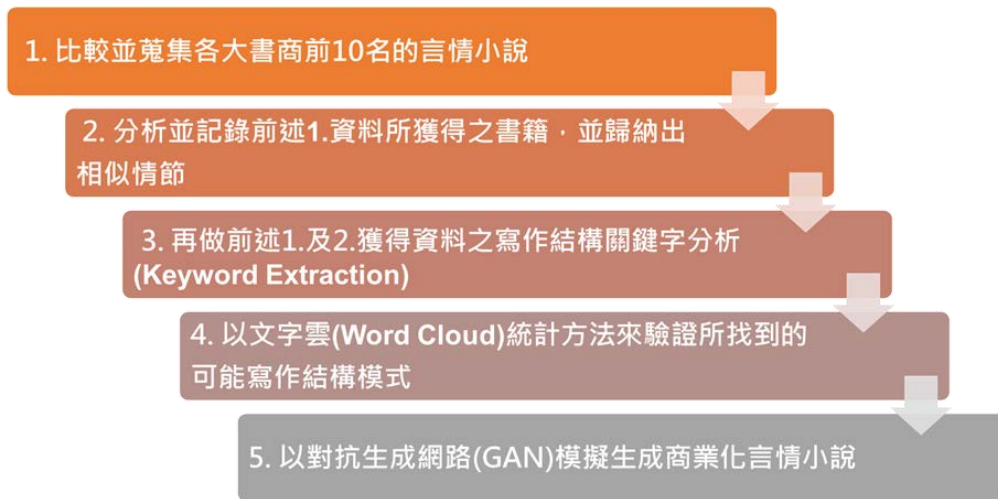


圖 1：研究流程圖

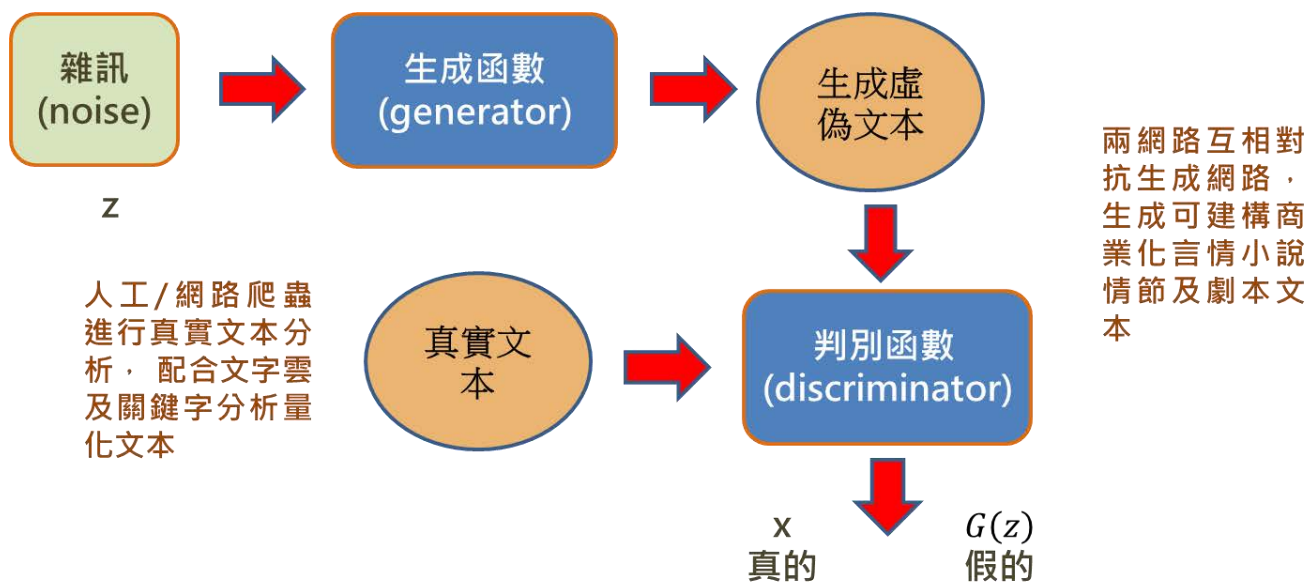


圖 2：生成對抗網路研究架構圖

參、文獻探討

言情小說定義：言情小說又稱愛情小說或稱為羅曼史小說，本質上是小說的一種類型，泛指以愛情故事為主體的小說，情節著重如何尋找一個理想的對象或維繫愛情關係[1]。由於愛情本身易引發動人的情緒，且小說劇情又常會利用人或物彼此的衝突情節來反映人物性格，從而成為小說作家常用的寫作題材。利用上述題材，再通過編輯完整的故事情節和具體結合特定時間空的環境，來描寫反映出主角對愛情的心理、狀態，以及符合世俗或通俗現實等社會生活的一種文學體裁[2][3]。

言情小說的現況及發展：受限於現代生活型態的改變，近代言情小說的主要讀者群為中學生，其次為家庭主婦。1990 年代為言情小說之全盛時期，至 2000 年後由於網路小說興起、讀者閱讀習慣轉變，加上盜版猖獗等因素，使得言情小說的市場逐漸由盛轉衰[31]。現今台灣出版社除了引入中國大陸網路連載小說外，也開始與 PUBU、讀墨、亂搭、Google 圖書等網路平臺合作，推出紙本書外的電子書版本[4]。目前言情小說的出版量每年仍高達一千多本（全台小說年出版量約四千本），佔全台小說出版量總量之 3 至 4 成[5]。因此對出版社而言，如何設計情節並找出暢銷言情小說的主要影響因素，實有其必要性。

文字雲及關鍵字分析：文字雲是一種分析關鍵字詞的視覺化統計描述技術，用於彙總用戶生成的文字標籤內容。因此其標籤一般會是獨立的詞彙，常常按筆劃、字母或數字進行順序排列，其重要程度又能通過改變字體大小或顏色來表現[6][7]。

生成對抗網路：生成對抗網路（Generative Adversarial Network，GAN）是一種非監督式學習（Unsupervised Learning）方法，該方法由伊恩·古德費洛(Ian J. Goodfellow)等人於 2014 年所提出，透過生成網路(Generator Network)與判別網路(Discriminator Network)相互博弈的方式進行學習[9]。生成網路從潛在空間中隨機取樣作為輸入，例如：言情小說的潛在空間有男同性戀者、女同性戀者及異性戀者，其輸出結果需要盡量模仿訓練集中的真實樣本。判別網路的輸入則為真實樣本及生成網路的輸出，其目的是將生成網路的輸出從真實樣本中盡可能分辨出來，例如：假設真實樣本大多數為異性戀者，生成網路則會輸出異性戀者。若此時真實樣本隨機取樣為男同性戀者，則判別網路輸出值將增大。而生成網路則要盡可能地欺騙判別網路，也就是使判別網路輸出值減小。兩個網路相互對抗、不斷調整參數，最終目的是使判別網路無法判斷生成網路的輸出結果是否真實[11]。

生成對抗網路最佳化目標函數：

$$\min_G \max_D V(D, G)$$

其中價值函數 $V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))]$ 、 $D(x)$ 為判別函數、 $G(z)$ 為生成函數、 $P_{data}(x)$ 為真實數據 x 服從之分佈，以及 $p_z(z)$ 為雜訊數據 z 服從之分佈。

上列公式是期望判別函數 D 儘可能增加 V (生成數據及真實數據的差異) 的值，生成函數 G 則儘可能

降低 V 的值，兩者反覆的互相對抗，最佳化彼此[24][25]。

例如，當校園愛情與都市愛情兩個數據庫在進行對抗生成時，數據庫內的字詞(校門、放學、書包……)將以出現頻率換算成機率的相對次數，也就是 $P_{data(x)}$ 。字詞機率比較大小，篩選出的出現頻率較高者，再將兩數據集合的差異最小化，也就是所謂的零和模式，當式(1)和式(2)不斷反覆運算直至結果幾乎不再變動。

判別函數：

$$\max_D \left(E_{x \sim P_{data(x)}} [\log D(x)] + E_{z \sim P_z(z)} [\log (1 - D(G(z)))] \right) \text{---(1)}$$

在實際操作中，我們需要先固定生成函數 G 以最佳化判別函數 D: 使上列公式的值增加，目標是**真實數據**(本研究的人工整理的關鍵字詞)期望被 D 分成 1，**生成數據**(本研究的電腦整理的關鍵字詞)期望被 D 分成 0，兩者以二分化分類。若前項(真實數據)被分錯，那麼會使 $\log (D(x)) \ll 0$ ，期望變成負無窮大。後項(生成數據)同理。當預期出現許多負無窮，表示可最佳化空間還很大。可以反覆修正參數，使 V 的數值增大。

生成函數：

$$\min_G \left(E_{z \sim P_z(z)} [\log (1 - D(G(z)))] \right) \text{---(2)}$$

在最佳化生成函數 G 的部分，則是希望 V 的值減小，讓判別函數 D 無法判斷真假數據。因為目標函數的首項不包含 G，屬於常數，因此可以忽略不計。由於生成函數 G 本身希望判別函數 D 在辨別時的值能夠越大越好(被 D 劃分 1 (真實數據))，因此可以推出上列兩行公式，兩者相等，但第二個公式的訓練效果較佳[24][25]。

文本生成 (Text Generation) 是透過機器學習技術結合自然語言處理(NLP)技術，使 AI 可具有人類程度的語言表達能力，能反應現代自然語言處理的發展狀態。GAN 結構整個模型只有兩個部分：1.生成函數 G；2.判別函數 D。生成函數 G 的目標是生成最接近於真實樣本的虛擬樣本，在以前沒有判別函數 D 的時候，生成器的訓練得依靠每次疊代來生成樣本，再與真實樣本的差異來進行修正。而判別函數 D 的出現改善了這一步驟，判別函數 D 的目標是儘可能準確地辨別生成樣本和真實樣本，此時生成函數 G 的訓練目標，就是最小化「生成-真實樣本差異」。為了儘量弱化判別函數 D 的辨別能力(此時訓練的目標函數中包含判別函數 D 的輸出)。因此無法直接分析非連續的函數結果(字詞)，只能以連續的函數結果(圖像數據)來分析與辨識。例如：將全文拍成圖進行圖像分類判別。

由於圖像數據在電腦中均可被表示為矩陣數值，因此圖像矩陣中的元素值是可微分的(為連續函數結果)，其數值可直接反映出圖像本身屬性特質，而這樣的組合就形成了圖像，也就是說，從圖像矩陣到圖像，不需要統計抽樣 (Sampling)，因此圖像可視為一連續函數，也就是其具有可微分的特性。字詞的文本數據是假設在特定的詞庫(可看成由一本書) 統計抽樣的關鍵字詞分析，所以預測下一個出現的字詞時，要先通過整理統計(例如:本研究的文字雲)由抽樣結果，去猜測未來可能出現或相關的字詞頻率進而推測出可能的小說情節[25]。

肆、研究過程與結果

研究資料來源

本研究資料採用台灣地區的四大網路書店公開資料進行相關分析，依序分別是**金石堂**、**博客來**、**墊腳石**與**MOMO**。

由於生成對抗網路可經由小量且真實的資料來產生大量的可供 AI 目標函數進行最佳化的演算方法，因此在文本創作上確實有一定的發展潛力，因此本研究利用該方法來進行相關的研究設計，研究架構如圖 3 所示。

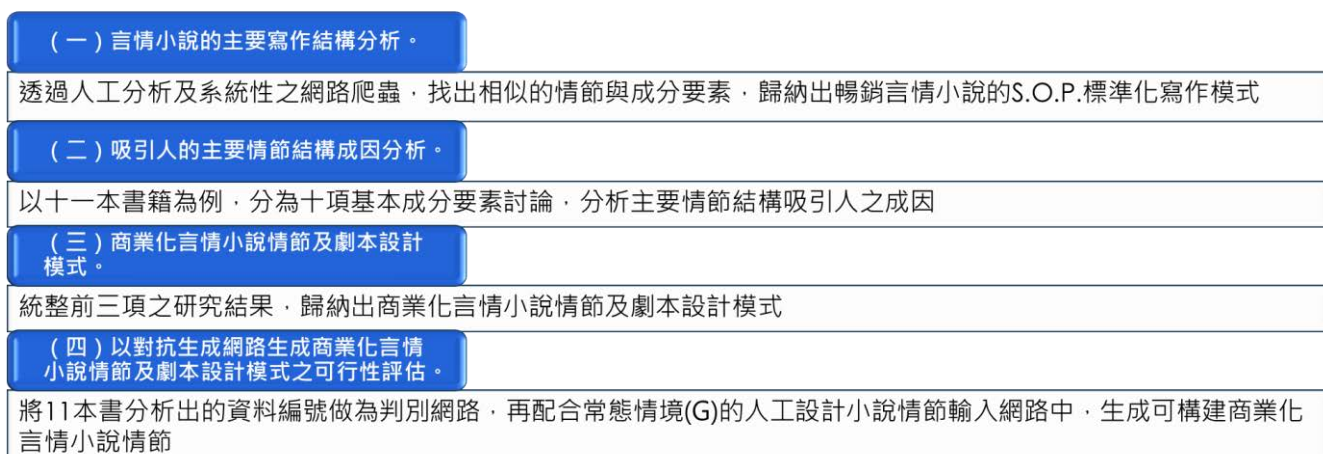


圖 3：研究架構圖(本研究團隊製作)

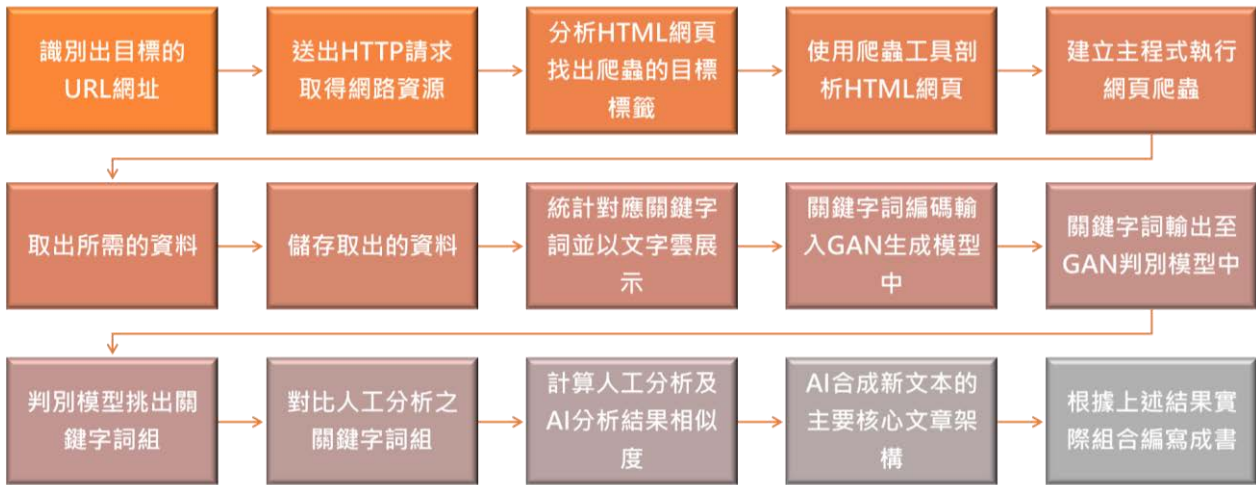


圖 4：爬蟲網路分析生成商業小說情節及劇本設計之實作架構圖(本研究團隊製作)

內容連載

放學鐘一敲，我迅速收拾好書包，準備到校門口跟柏軒會合。

下了樓梯，看到站在樓梯口的人，我愕然的問：「梁少桓，你站在這裡做什麼？」

「等妳啊，跟妳一起去校門口。」

「為什麼？」我吃驚。

「為了親眼看妳被甩呀！」他哈哈笑，邁開步伐往前走。

我馬上衝去拉住他：「等一下，哪有這樣的？」

「大家都這麼熟了，有什麼關係？柏軒不會介意啦，而且我哪知道妳會不會騙我？」

我又氣又急，阻止不了他，最後只能擋在他前面：「梁少桓，你不要鬧了啦！」

「妳很煩耶，到底要不要過去啊？」他淡淡瞥我一眼，再遠望我身後的方向，「他到了，妳動作快點。」

我回頭看見柏軒已經在校門口，心跳倏地漏跳一拍。

我渾身僵硬，跟著梁少桓的腳步走向他，胸口的壓迫感越來越重，重到我還沒走到那裡，腳卻已經慢慢停下來，最後整個人站在原地，無法動彈。

圖 5：爬蟲協助生成關鍵字詞頻率分析元素示意圖(本研究團隊製作)

本研究以金石堂、博客來、墊腳石、MOMO 等四大暢銷網路書店為研究對象，此四家網路書店之所以暢銷的主要因為：具有知名度、客群來源廣泛，且暢銷排行榜資料完整。依各書店的年度十大暢銷言情小說，分別進行網路聲量分析，做出網路聲量圖(圖 6)。利用文字雲及關鍵字分析，四大

書店暢銷言情小說之個別文字雲圖(圖 7)及合併的文字雲圖(圖 8)。

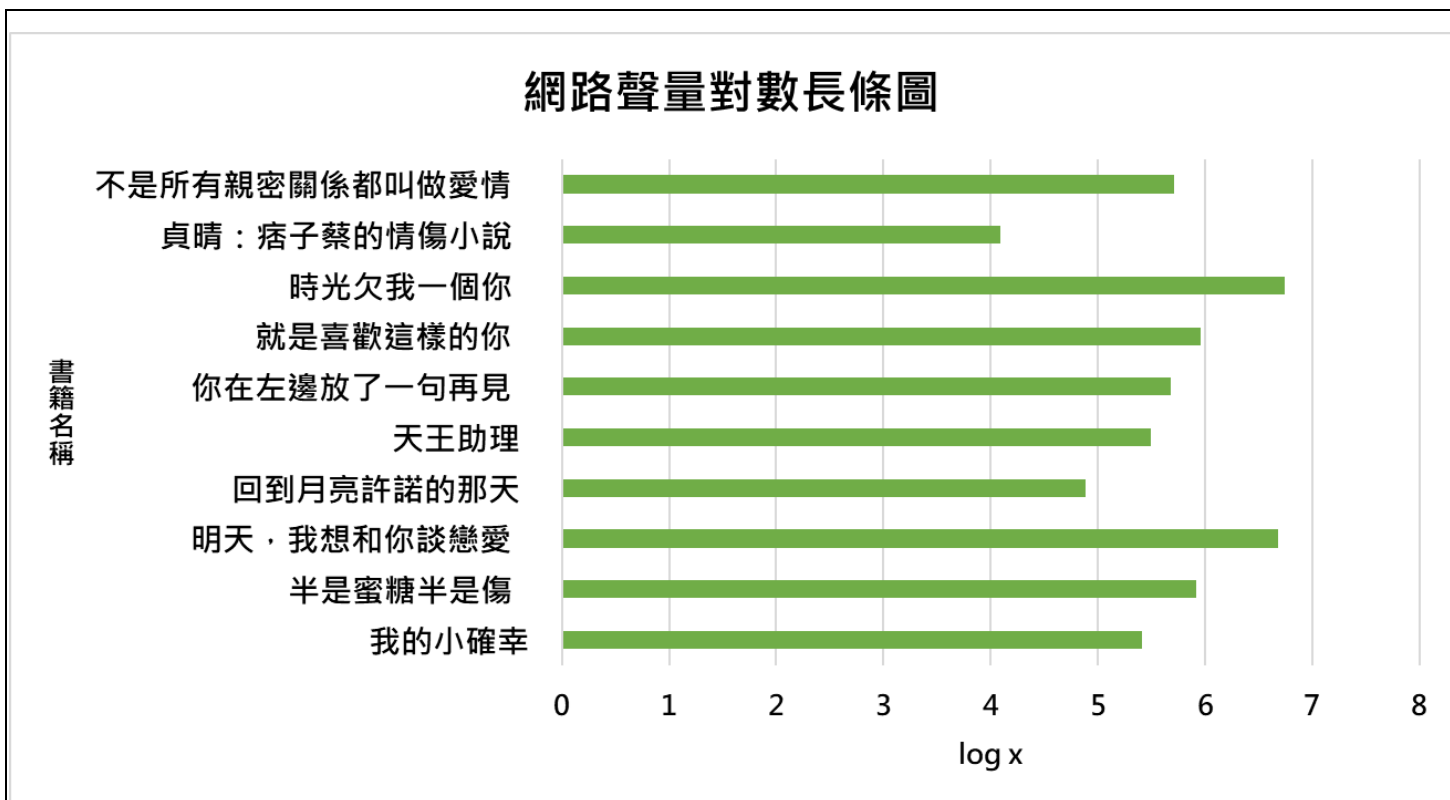


圖 6-1：金石堂十大暢銷言情小說網路聲量圖

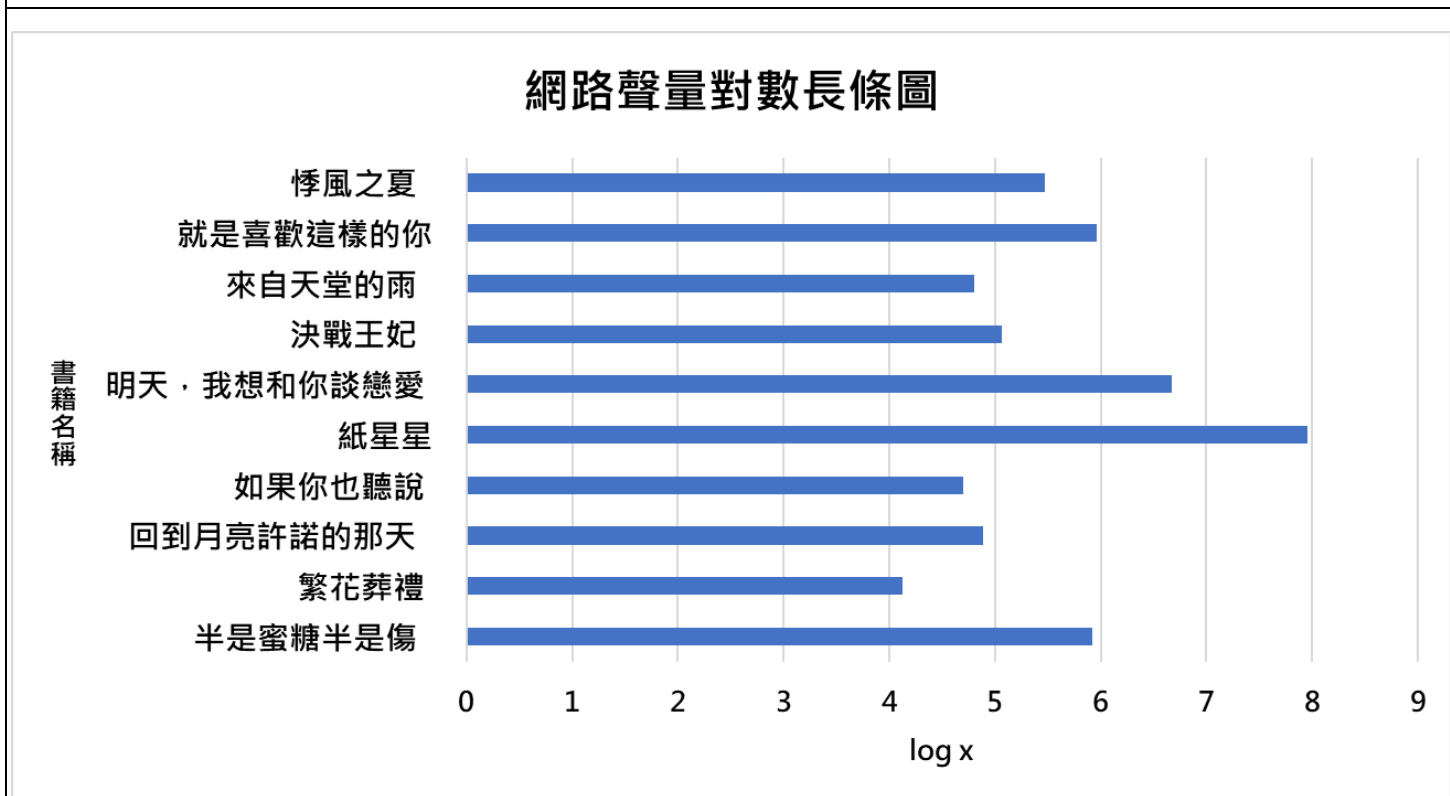


圖 6-2：博客來十大暢銷言情小說網路聲量圖

網路聲量對數長條圖

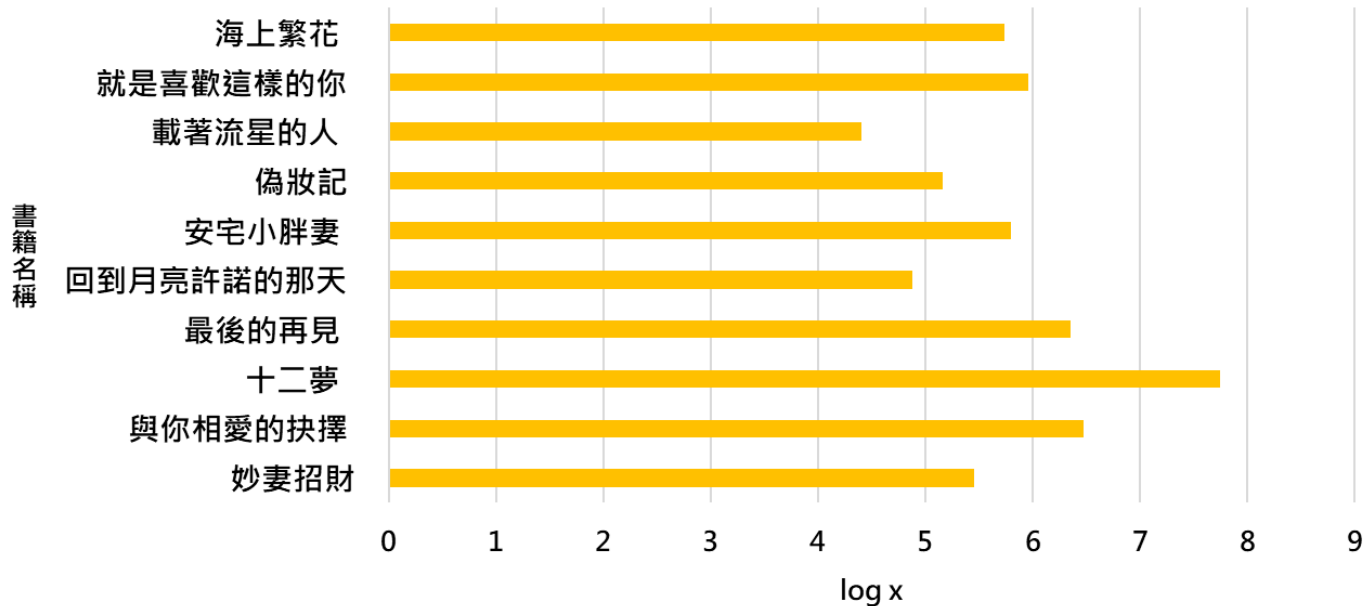


圖 6-3：墊腳石十大暢銷言情小說網路聲量圖

網路聲量對數長條圖

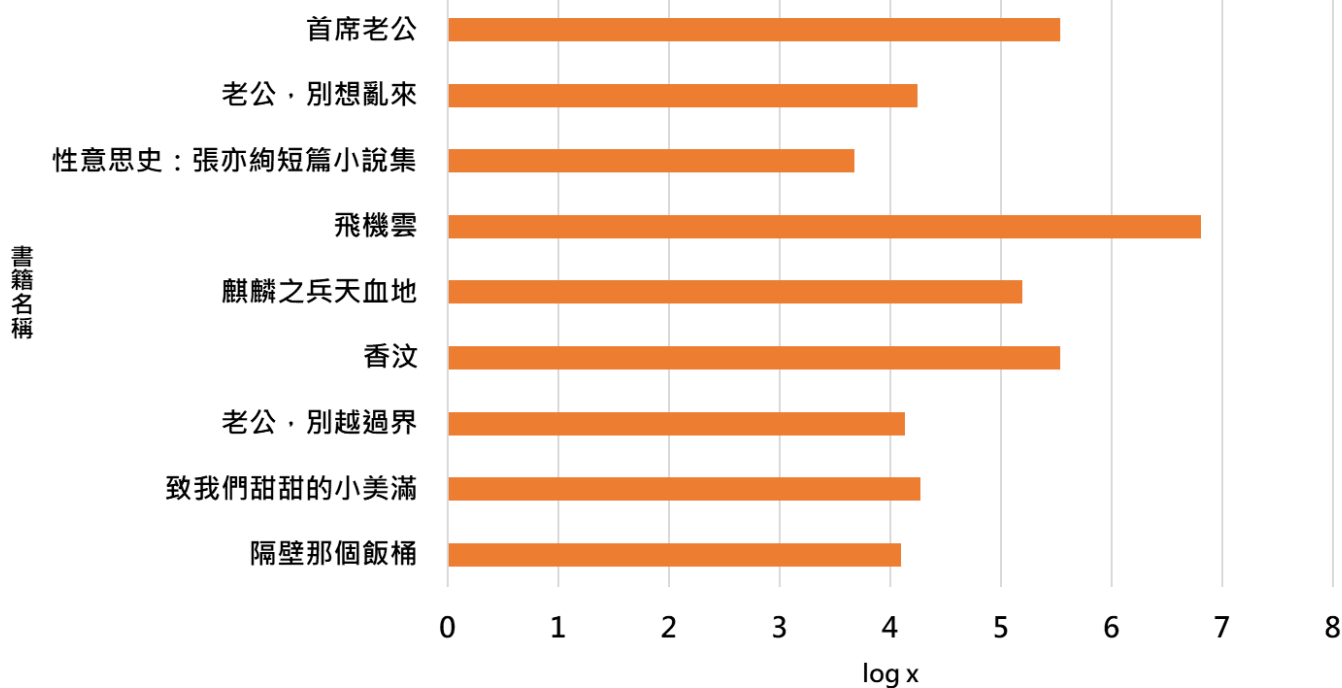


圖 6-4：MOMO 十大暢銷言情小說網路聲量圖

圖 6:四大書店暢銷言情小說及網路聲量分析(本研究團隊製作)



圖 7-1：金石堂網路聲量文字雲圖



圖 7-2：博客來網路聲量文字雲圖



圖 7-3：墊腳石網路聲量文字雲圖



圖 7-4：MOMO 網路聲量文字雲圖

圖 7:四大書店暢銷言情小說各自的文字雲分析圖 (本研究團隊製作)

表 1：暢銷言情小說的寫作成分要素概念表(本研究團隊製作)

故事背景(x1)	相識方式(x2)	交往過程(x3)	結局(x4)	故事走向(x5)
男主角特質(x6)	女主角特質(x7)	探討議題(x8)	男主角是否初戀 (x9)	女主角是否初戀 (x10)

在公式(1)及公式(2)的公式彼此反覆疊代運算執行下，我們可以最佳化出生成函數 G 的最佳解：生成函數 G 產生的機率分配要和真實機率分配一致，使判別函數 D 無法判斷出生成數據與真實數據，即： $P_{data(x)} = P_g(x)$ 在此條件下，代表已經知道是最佳解了，也就是生成函數 G 產生出來的機率分配會和真實機率分配一致 [24] [25]。因此也表示電腦可以提供作者或出版社要的文本生成關鍵資料。

生成商業化言情小說情節程式撰寫 AI 模擬：

```

conda -V (偵測版本)
conda env list (列出環境)
conda update conda(版本升級)
conda create --name myenv python=3.7(創造一個命名為-python3.7的環境)
conda activate py37(啟動環境)
>conda install -c anaconda ipykernel
python -m ipykernel install --user --name=py37
    
```

圖 9：python 3.7 程式模擬

步驟 1:使用 anaconda 中的 jupyter 的 python 3.7 編輯器以撰寫程式

```

out[0]: array([ 3,  5,  7,  9, 11])
In [ ]: array2 * array
In [ ]: array[0]
In [ ]: array[3]
In [ ]: array
In [ ]: array.shape
In [ ]: tang_list = [1,2,3,4,5]
          tang_list.shape
In [ ]: np.array([[1,2,3],[4,5,6]])
In [13]: a=5
          b=7
          c=a*b
          print(c)
          12
In [14]: c=a*b
          print(c)
          35
    
```

圖 10：jupyter 程式模擬

步驟 2:連線至國家網路中心執行相關程式進行以程式組合詞組產出短文

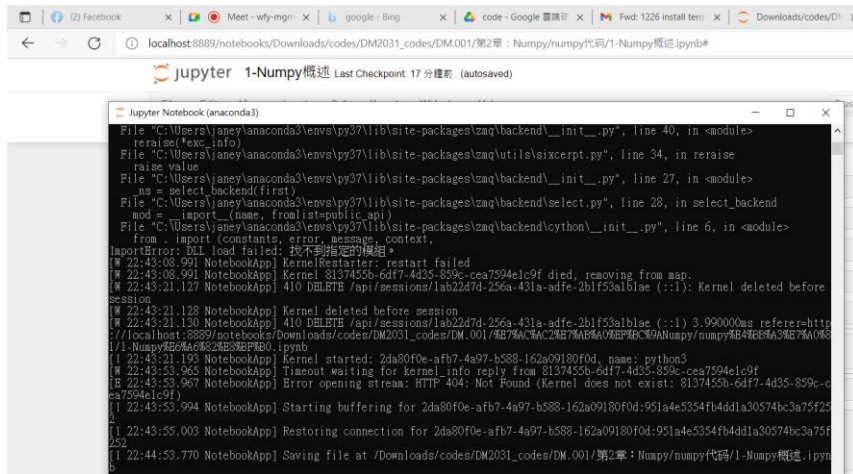


圖 11：國家網路中心遠端網路連線中

步驟 3: 遠端執行程式過程中並進行 DEBUG，查找錯誤來源並進行文本生成系統模擬。

```
In [2]: corpus = [
    '故事發展',
    '相識方式',
    '交際過程',
    '結局',
    '男主角特質',
    '女主角特質',
    '探討議題'
]
labels = ['校園愛情', '互看不順眼', '日久生情', '喜劇', '安靜沉穩', '活潑開朗', '學習諒解', '霸氣中住', '溫馨可靠', '優秀放蕩', '學醫諒解', '青少年成長', '初戀']
corpus = np.array(corpus)
corpus_df = pd.DataFrame({'寫作要素': corpus,
                          '建議模式': labels})
corpus_df = corpus_df[['寫作要素', '建議模式']]
corpus_df
```

圖 12：國家網路中心遠端網路連線結果資料運算及設定回傳(本研究實驗結果)

步驟 4:輸出 AI 生成參考文本組合詞組之短文。

對抗生成網路建議模式	網路爬蟲對應之內容關鍵字
校園愛情	放學、校門、書包、課本、學長、學姊.....
互看不順眼	殺人、氣、急、不耐、不理不睬、不悅.....
日久生情	發現喜歡上、漸漸喜歡上、陪伴、改觀.....
喜劇	男朋友、女朋友、我也喜歡你、在一起.....
安靜沉穩	靦腆、靜靜陪著、面無表情.....
活潑開朗	燦爛笑容、開朗、活力滿滿、沒心沒肺.....
學習諒解	和好、和解、緩和、不再排斥、體諒.....

圖 13：網路爬蟲內容關鍵字資料庫

伍、討論

(一)言情小說的主要寫作組成結構

根據上述所蒐集到的資料，挑選十一本書做為 $P_{data(x)}$ 進行研究[12-22]。本研究參考文獻[1-5]的分析，分別設計出以背景、初識、過程、結局、男、女主角特質、探討議題、故事走向及男、女主角是否為初戀等十項言情小說基本要素，本研究依序編為變數 $x1\sim x10$ ，為人氣暢銷言情小說的主成份進行統整。透過人工分析及系統性之網路爬蟲[23]兩項方法，進而找出相似的情節與成分要素，歸納出暢銷言情小說的 S.O.P.標準化寫作模式，表 2 至表 5。

表 2: 暢銷言情小說的寫作成分要素模式分析之一 (本研究團隊製作)

書名	故事背景(x1)	初次相識(x2)	交往過程(x3)
回到月亮許諾的那天	校園愛情(x11)	互看不順眼(x21)	日久生情
最後的再見	校園→都市愛情 (x12)	互看不順眼	日久生情
載著流星的人	校園→都市愛情	重組家庭的兄妹	日久生情
紙星星	校園愛情	互看不順眼	日久生情
首席老公	都市愛情(x13)	互看不順眼	日久生情
來自天堂的雨	校園愛情	女主角的朋友牽線	日久生情
與你相愛的抉擇	校園愛情	男主角注意女主角	日久生情
老公，別越過界	校園→都市愛情	社團認識	舊情復燃
如果你也聽說	校園愛情	男主角注意女主角	相知相惜
天王助理	都市愛情	男主角注意女主角	日久生情
十二夢	校園愛情	男主角愛慕女主角	日久生情

表 3: 暢銷言情小說的寫作成分要素模式分析之二(本研究團隊製作)

書名	結局(x4)	故事走向(x5)
回到月亮許諾的那天	男主角與女主角在一起	喜劇
最後的再見	男主角與女主角沒有在一起	悲劇
載著流星的人	男主角與女主角在一起	喜劇
紙星星	男主角與女主角沒有在一起	悲劇
首席老公	男主角與女主角在一起	喜劇
來自天堂的雨	男主角與女主角在一起	喜劇
與你相愛的抉擇	男主角與女主角在一起	喜劇
老公，別越過界	男主角與女主角在一起	喜劇
如果你也聽說	男主角與女主角沒有在一起	悲劇
天王助理	男主角與女主角在一起	喜劇
十二夢	男主角與女主角沒有在一起	悲劇

研究結果顯示，言情小說主要暢銷模式可分為三大類故事背景，校園、都市以及由校園走向都市。男女主角的初次相識方面則顯示**互看不順眼**的情節比例較高。在交往過程中高機率不是一見鍾情，而是**日久生情**。故事走向雖有悲有喜，但**喜劇**仍佔多數比例。

表 4: 暢銷言情小說的寫作成分要素模式分析之三 (本研究團隊製作)

書名	男主角特質(x6)	女主角特質(x7)
回到月亮許諾的那天	霸氣，害羞	富正義感
最後的再見	安靜沉穩，壓抑，霸氣，優秀帥氣	霸氣中性，個性衝動倔強，開朗活潑，直來直往
載著流星的人	安靜沈穩，貼心可靠，善於傾聽	運動健將，個性堅強倔強，直來直往
紙星星	話少神秘，兇狠，成熟穩重，叛逆，頭痛人物	優秀執著，懦弱聽話，功課優秀，多才多藝，模範生

首席老公	成熟穩重，優秀帥氣，多金總裁，溫柔可靠	外貌亮麗，優秀，善解人意，千金小姐，開朗賢淑，討人喜愛
來自天堂的雨	安靜沈穩、游泳校隊、校草	活潑開朗，害怕背叛，勇於面對
與你相愛的抉擇	安靜沉穩，成熟穩重，溫柔可靠，優秀帥氣	知性，個性堅強，性格善良，文靜懂事，平凡膽小
老公，別越過界	成熟穩重，優秀帥氣，多金總裁	溫柔優雅，外貌亮麗，優秀，善解人意
如果你也聽說	口吃，膽小懦弱，細心體貼，不善交際	安靜沉穩，討人喜愛，功課優秀，外貌亮麗，多才多藝，模範生
天王助理	開朗活潑，善良暖男，帥氣多金，偶像歌手	優雅知性，外貌亮麗，功課優秀，模範生，情商低
十二夢	膽小懦弱，畏畏縮縮，不敢面對	完美，功課優秀，性格善良，外貌亮麗，模範生

表 5:言情小說基本架構分析之四 (本研究團隊製作)

書名	探討議題(x8)	男主角是否為初戀(x9)	女主角是否為初戀(x10)
回到月亮許諾的那天	學習諒解，青少年成長，校園霸凌	是	是
最後的再見	把握當下，青少年成長，生死議題，夢想，穿越重生	是	是
載著流星的人	重組家庭，再婚議題，手足情誼，青少年成長	否	否
紙星星	青少年成長，手足情誼，青少年犯罪，心理疾病，家庭問題，性別議題，夢想	是	是

首席老公	家庭問題，幼兒教育，婚姻議題，夢想	是	是
來自天堂的雨	珍惜生命，校園霸凌，青少年成長	是	否
與你相愛的抉擇	青少年成長，校園霸凌，心理疾病，家庭問題，性別議題，珍惜生命，穿越重生	否	是
老公，別越過界	家庭問題，幼兒教育，婚姻議題	是	是
如果你也聽說	青少年成長，家庭問題，精神虐待，校園霸凌，手足情誼	是	是
天王助理	心理疾病，夢想，網路暴力	是	否
十二夢	青少年成長，校園霸凌，心理疾病，家庭問題，多重人格，校園性騷擾	是	是

(二)吸引人的主要情節結構及成因

由表 1 至表 5 內容得知，暢銷小說的男主角多具備**優秀帥氣、安靜沉穩、成熟穩重、溫柔可靠**等特質，女主角則多數具有**功課優秀，外貌亮麗，善良、開朗**等特質，男、女主角也多是初戀。而在研究過程中我們也發現當**探討議題越多元**時，書籍網路聲量也越高。根據此結果我們推測當探討議題越多元，越能增加讀者的認同感，並吸引不同類型的讀者，從而引起讀者共鳴並以行動支持書籍，增加書籍銷售量及網路聲量。

(三)商業化言情小說情節及劇本設計模式

研究結果顯示，不同寫作計畫、轉譯及回顧各階段的認知行為並無太大差異。他們都很少在寫作前設定寫作目標或構思段落，不過作者多半會參考時代背景來進行寫作，探討多元化的社會議題，以期待增加讀者的認同感，因此以商業化言情小說情節需兼顧多項因素。以本次分析的**紙星星**而言，它是整體排序第一的小說，也探討了多項議題：**青少年成長，手足情誼，青少年犯罪，心理疾病，家庭問題，性別議題，夢想**使讀者易幻想成為書中的男、女主角，也使他們易聯想到自己的生活環境最終沉醉書中情節。值得注意的是網際網路普及後，開始出現網路作家進行寫作網路小說，因此有不少作品增加對網戀的描寫，其目標皆是令讀者產生共鳴。

(四)以對抗生成網路生成商業化言情小說情節及劇本設計模式之可行性評估

Out[2]:

	寫作要素	建議模式
0	故事背景	校園愛情
1	相識方式	互看不順眼
2	交往過程	日久生情
3	結局	喜劇
4	男主角特質	安靜沉穩，溫柔可靠，優秀帥氣
5	女主角特質	霸氣中性，活潑開朗
6	探討議題	學習諒解，青少年成長，初戀

圖 14：國家網路中心遠端計算回傳建議文本組合

電腦分析了上述 11 本書及 10 大因素之後，建議寫作模式：**故事背景**以校園愛情為主、男女主角**相識方式**先互看不順眼再由男生開始喜歡女生、**交往過程**是因日久生情、最後結局男主角與女主角以**喜劇**的方式**在一起**、**男主角特質**具備安靜沉穩，溫柔可靠，優秀帥氣且**女主角特質**以霸氣中性，開朗活潑、**探討議題**為學習諒解，青少年成長、**男主角女主角都是初戀**。

寫作要素	人工分析結果	AI分析結果
故事背景	校園愛情	校園愛情
相識方式	互看不順眼	互看不順眼
交往過程	日久生情	日久生情
結局	喜劇	喜劇
男主角特質	成熟穩重、安靜沉穩、溫柔可靠、優秀帥氣	安靜沉穩、溫柔可靠、優秀帥氣
女主角特質	外貌亮麗、功課優秀、霸氣中性、活潑開朗	霸氣中性、活潑開朗
探討議題	學習諒解、青少年成長、手足情誼、夢想	學習諒解、青少年成長、初戀

圖 15：人工分析及 AI 分析結果對比(本研究實驗結果)

陸、結論

本研究成功地完成四項研究目標，也探討了不同寫作表現的暢銷言情小說模式及寫作內容結構加以分析，同時將其轉化為對抗生成網路生成寫作模式之可行性分析。

在研究過程中，我們發現在文章完成後，作者常回顧全文，且在修改作品時，喜歡把對應的歷史及時事予以改編到劇情中，再對文章整體結構進行修正。由結果看來，無論是何種型態的言情小說，終極目標就是希望讀者幻想自己成為男、女主角，使其懷抱少女情懷與戀愛夢想。特別是以言情小說的讀者為對象的內容[4]，他們多半以純愛為主題，主角個性多屬較單純的一類，且常以初戀為題材，過程也多是專一、簡單的愛情。主角通常較為優秀、外貌亮麗、善良、成熟或男、女主角互相暗戀對方但長久不告白的情節。

而關於生成對抗網路的技術，目前多被應用在圖像與影音的生成、合成、辨識、修復等等資料方面，更進階的則是輸入文本描述，便能生成與文本描述形容相符的圖像等，因此運算時須使用大量的系統資源，但最終可利用對抗生成網路來改善影視文化出版業的困擾。而由於生成對抗網路作為非監督學習的重要訓練方法，它具有訓練資料不需要事先以人力處理標籤，機器面對資料時，做的處理是依照關聯性去歸類、進而找出潛在規則與套路形成集群，不對資訊有正確或不正確的判別，因此必須事前提供良好訓練文本資料，故未來應全面以網路爬蟲的方式，亦即以爬蟲軟體更新不同小說的內容，之後再以關鍵字分析與文字雲量測出最佳文本資料，來強化判別網路部分的資訊，具體提升非監督學習的生成對抗網路分析計算的精確程度。

柒、未來展望

隨著科技發展，相信往後由人工智慧大量生成小說的時代即將到來。期許此項研究成果能被廣泛應用於文本生成方面，具體降低各類型小說作家及編劇的工作壓力，減少更多的人力及時間成本。因此，相信本研究所提出的言情小說情節及劇本設計生成資料架構之可行性報告，可確實具體降低未來影視文化出版業的相關成本。另外，亦期許未來可將此技術應用於國外知名小說，使臺灣出品的小說不僅僅能暢銷國內，更能推廣至國際，但仍需考慮各國文化及背景的差異性，例如：在回教國家中，盛行的言情小說模式極有可能非一夫一妻制，而是一夫多妻制。此項議題將是未來 AI 技術應用於國外小說的撰寫時，具挑戰性的問題之一。

捌、參考文獻

- [1] 李孟純、陳心如、李品蓁。(2016)。走訪愛情香格里拉—言情小說之人物情節與讀者心理初探，中學生小論文。
- [2] 楊若慈。(2015)。那些年，我們愛的步步驚心-台灣言情小說浪潮中的性別政治。台北市：秀威資訊科技股份有限公司。
- [3] 藝術與建築索引典—言情小說。(2011)。資料取自：<https://zh.wikipedia.org/wiki/%E6%84%9B%E6%83%85%E5%B0%8F%E8%AA%AA>。
- [4] 謝奇任。(2016)。致我們的青春：臺灣、日本、韓國與中國大陸的網路小說產業發展。臺北市：威秀經典。
- [5] 倪采青。(2010)。言情小說 編織另類作家夢，Career 職場情報誌 9 月號第 413 期。
- [6] Martin Halvey and Mark T. Keane (2007), An Assessment of Tag Presentation Techniques, poster presentation at WWW.
- [7] 趙維孝。(2020)如何快速掌握社群議題趨勢 – 從文字雲解析 4 個月東京奧運聲量！。資料取自：<https://blog.qsearch.cc/2020/03/tokyo2020/>
- [8] 陸怡琮、曾慧禎。(2004)。不同寫作表現的國小六年級學童在寫作歷程中的後設認知行為之比較，國北師學報 17 卷第 2 期。
- [9] Goodfellow, Ian J.; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua. (2014). Generative Adversarial Networks.
- [10] Caesar, Holger. (2019). A list of papers on Generative Adversarial (Neural) Networks: [nightrome/really-awesome-gan](https://github.com/holgercaesar/really-awesome-gan).
- [11] Jakub Langr, Vladimir Bok。(2020)。GAN 對抗式生成網路，旗標。
- [12] Misa(2020)。回到月亮許諾的那天。台北市：城邦文化。
- [13] 準擬佳期(2020)。天王助理。台北市：高寶國際。
- [14] 晨羽(2019)。如果你也聽說。台北市：城邦文化。
- [15] 晨羽(2014)。紙星星。台北市：城邦文化。
- [16] 晨羽(2014)。來自天堂的雨。台北市：城邦文化。

- [17] 琉影(2020)。與你相愛的抉擇。台北市：城邦文化。
- [18] 晨羽(2015)。十二夢。台北市：城邦文化。
- [19] 瑪琪朵(2019)。最後的再見。台北市：城邦文化。
- [20] 晨羽(2013)。載著流星的人。台北市：城邦文化。
- [21] 桑蕾拉(2020)。老公，別越過界。台北市：狗屋。
- [22] 夏洛蔓(2020)。首席老公。台北市：狗屋。
- [23] 洪錦魁。(2021)。Python 網路爬蟲：大數據擷取、清洗、儲存與分析：王者歸來，深智數位。
- [24] GAN 入門理解及公式推導 資料取自：<https://zhuanlan.zhihu.com/p/28853704>
- [25] 機器不學習：強化學習在生成對抗網絡文本生成中扮演的角色。
資料取自：<https://kknews.cc/code/v634p4q.html><https://kknews.cc/zh-tw/code/v634p4q.html>
- [26] 使用 GPT2-Chinese 生成中文小說。
資料取自：https://www.cc.ntu.edu.tw/chinese/epaper/0058/20210920_5808.html
- [27] [NLP] 使用 aitextgen 套件來生成文本。
資料取自：<https://clay-atlas.com/blog/2021/10/13/nlp-cn-python-aitextgen-generate-text/>
- [28] 我用 OpenAI 文本生成器續寫了《復仇者聯盟》。
資料取自：<https://www.inside.com.tw/article/16479-open-ai-edit-the-avengers>
- [29] Md. Akmal Haidar and Mehdi Rezagholizad. (2019). TextKD-GAN: Text Generation using Knowledge Distillation and Generative Adversarial Networks. Advances in Artificial Intelligence (pp.107-118)
- [30] Mahmoud Hossam¹, Trung Le¹, Michael Papisimeon², Viet Huynh¹, Dinh Phung. (2021).Text Generation with Deep Variational GAN. 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.
- [31] 陳俞辰(2021)。由人工智慧製作出吸引青少年的言情小說之可行性研究。中學生小論文。資料取自：<https://www.shs.edu.tw/works/essay/2021/03/2021031417470854.pdf>

【評語】 052702

這個研究使用文字雲與對抗生成網路分析，試圖找到暢銷言情小說的寫作架構，以及有哪些主角本身或主角間互動的特性，以及故事發展的議題。似乎是希望能夠做到自動生成夠吸引人的言情小說。提供建議如下：

1. 研究成果的呈現，主要是呈現資料的分佈。但作品的主要假說不明確，想要解決什麼問題也未清楚陳述。
2. 作者從自己挑選的言情小說中所抽取出來的片段或是語句是否真能代表所有這類小說會出現的詞彙？這點作者未加以驗證。
3. 抽取出的特性，如何最佳化，以生成商業化喜說的情節與架構，未見可行性的評估。
4. 作家的情緒、心境、以及想法、思緒等都是影響到一部小說是否成功的因素。由 GAN 所創作出來的小說是否真能達成或表現出這些內心或是心理層次的意境？建議可以多加討論。
5. 建議可增加暢銷與非暢銷，或言情與非言情小說的比較。

作品簡報

由人工智慧生成吸引人的言情小說之 可行性研究

科 別：行為與社會科學科

組 別：高中職組

編 號：

關鍵字：對抗生成網路、網路爬蟲、言情小說

壹、研究動機

- 希望了解為何言情小說能在眾多類型的書籍中吸引我們身邊的同儕。
- 結合現今的人工智慧網路設計出可以供電腦自動生成暢銷的言情小說寫作題材。

貳、研究目的、流程及架構

• 研究目的

1. 言情小說的主要寫作結構分析。
2. 吸引人的主要情節結構成因分析。
3. 歸納商業化言情小說情節及劇本設計模式。
4. 以對抗生成網路生成商業化言情小說情節及劇本設計模式之可行性評估。

• 研究流程

1. 比較並蒐集各大書商前10名的言情小說
2. 分析並記錄前述1.資料所獲得之書籍，並歸納出相似情節
3. 再做前述1.及2.獲得資料之寫作結構關鍵字分析 (Keyword Extraction)
4. 以文字雲 (Word Cloud) 統計方法來驗證所找到的可能寫作結構模式
5. 以對抗生成網路 (GAN) 模擬生成商業化言情小說

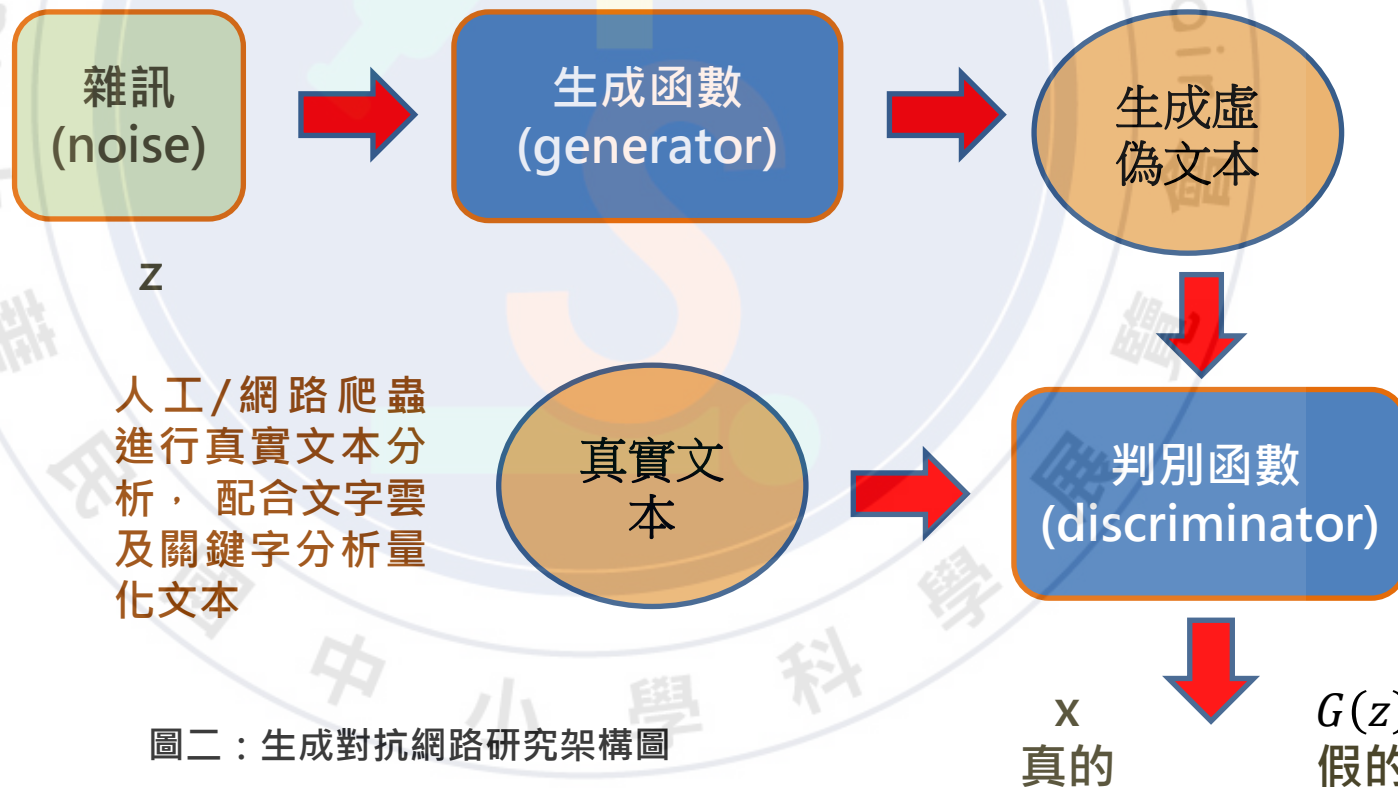
圖一：研究流程圖

貳、研究目的、流程及架構

- 研究架構：

生成對抗網路(Generative Adversarial Network)

- 「對抗學習」機制
- 自動最佳化預測能力
- 架構設計：



兩網路互相對抗生成網路，生成可建構商業化言情小說情節及劇本文本

圖二：生成對抗網路研究架構圖

參、文獻探討

- **言情小說**：又稱愛情小說或稱為羅曼史小說，本質上是小說的一種類型，泛指以愛情故事為主體的小說，情節著重如何尋找一個理想的對象或維繫愛情關係。
- **文字雲及關鍵字分析**：文字雲是一種分析關鍵字詞的視覺化統計描述技術，用於彙總用戶生成的文字標籤內容。
- **生成對抗網路**：生成對抗網路 (Generative Adversarial Network , GAN) 是一種非監督式學習 (Unsupervised Learning) 方法，該方法由伊恩·古德費洛 (Ian J. Goodfellow) 等人於2014年所提出，透過生成網路 (Generator Network) 與判別網路 (Discriminator Network) 相互博弈的方式進行學習 [9]。

生成對抗網路最佳化目標函數：

$$\min_G \max_D V(D, G)$$
$$V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} \left[\log \left(1 - D(G(z)) \right) \right]$$

- $D(x)$ 為判別函數
- $G(z)$ 為生成函數
- $P_{data}(x)$ 為真實數據 x 服從之分佈
- $p_z(z)$ 為雜訊數據 z 服從之分佈

肆、研究過程與結果

- STEP 1. 網路爬蟲進行文本分析
- STEP 2. 由文本分析出關鍵字詞
- STEP 3. 由關鍵字詞整理主要情節



圖三：四大書商暢銷言情小說綜合聲量總排名文字雲圖

內容連載

放學鐘一敲，我迅速收拾好書包，準備到校門口跟柏軒會合。

下了樓梯，看到站在樓梯口的人，我愕然的問：「梁少桓，你站在這裡做什麼？」

「等妳啊，跟妳一起去校門口。」

「為什麼？」我吃驚。

「為了親眼看妳被甩呀！」他哈哈笑，邁開步伐往前走。

我馬上衝去拉住他：「等一下，哪有這樣的？」

「大家都這麼熟了，有什麼關係？柏軒不會介意啦，而且我哪知道妳會不會騙我？」

我又氣又急，阻止不了他，最後只能擋在他前面：「梁少桓，你不要鬧了啦！」

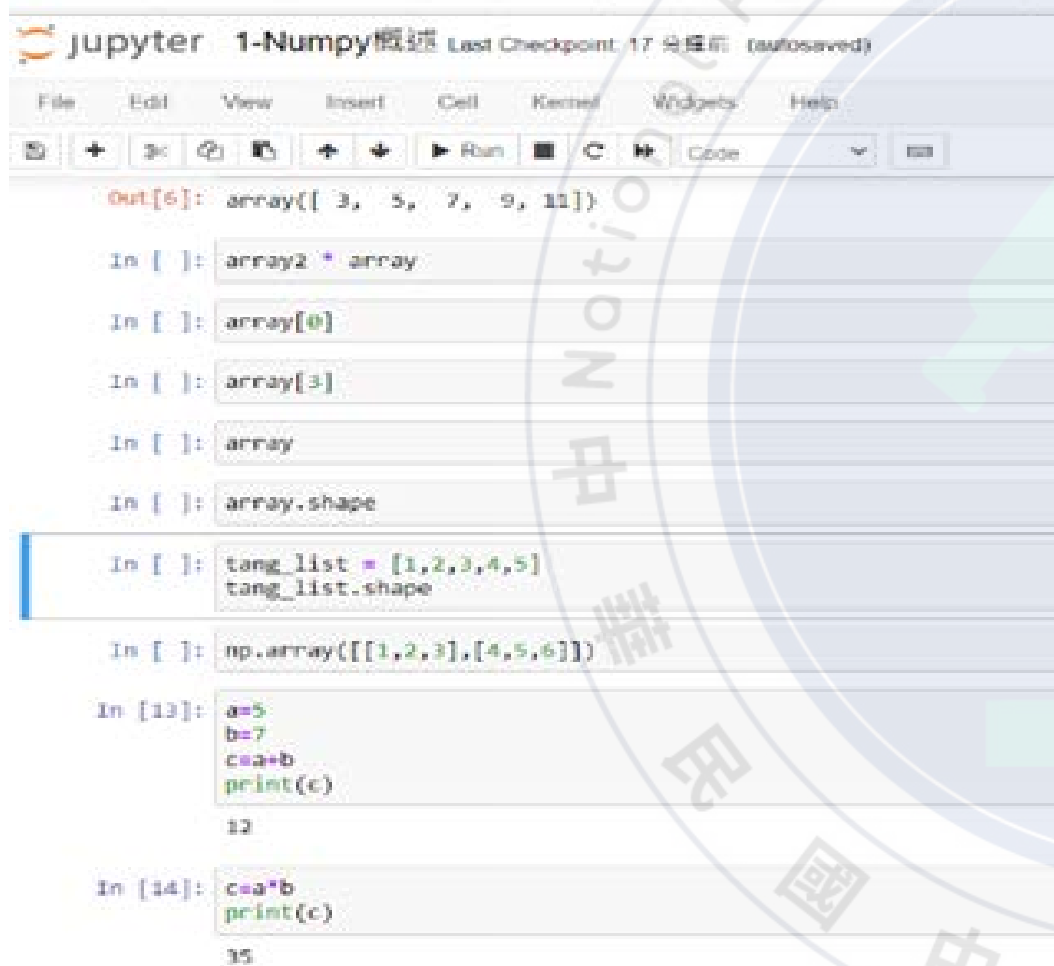
「妳很煩耶，到底要不要過去啊？」他淡淡瞥我一眼，再遠望我身後的方向，「他到了，妳動作快點。」

我回頭看見柏軒已經在校門口，心跳倏地漏跳一拍。

圖四：網路爬蟲協助對抗生成網路依照字詞出現頻率歸納情節元素之示意圖

肆、研究過程與結果

- 遠端連線至國網中心，查找錯誤來源並進行文本生成AI系統模擬
- 輸出分析結果，生成參考文本及組合詞組之短文



```
jupyter 1-Numpy 概覽 Last Checkpoint: 17 分鐘前 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
+ -> <-> Run C Code
Out[5]: array([ 3,  5,  7,  9, 11])
In [ ]: array2 * array
In [ ]: array[0]
In [ ]: array[3]
In [ ]: array
In [ ]: array.shape
In [ ]: tang_list = [1,2,3,4,5]
          tang_list.shape
In [ ]: np.array([[1,2,3],[4,5,6]])
In [13]: a=5
          b=7
          c=a+b
          print(c)
          12
In [14]: c=a*b
          print(c)
          35
```

圖五：jupyter程式模擬



```
In [2]: corpus = ['故事類型',
                  '閱讀方式',
                  '交往過程',
                  '結局',
                  '男主角特質',
                  '女主角特質',
                  '探討議題']
labels = ['校園愛情', '互看不順眼', '日久生情', '喜劇', '安靜沉穩, 溫柔可靠, 優秀帥氣', '霸氣中性, 活潑開朗', '學習諳解, 青少年成長, 初戀']
corpus = np.array(corpus)
corpus_df = pd.DataFrame({'寫作要素': corpus,
                          '建議模式': labels})
corpus_df = corpus_df[['寫作要素', '建議模式']]
corpus_df
```

圖六：國家網路中心遠端網路連線結果資料運算及設定回傳

伍、討論

Out[2]:

寫作要素	建議模式
0 故事背景	校園愛情
1 相識方式	互看不順眼
2 交往過程	日久生情
3 結局	喜劇
4 男主角特質	安靜沉穩，溫柔可靠，優秀帥氣
5 女主角特質	霸氣中性，活潑開朗
6 探討議題	學習諒解，書少生成長，初戀

- 國家網路中心遠端計算回傳建議文本組合

對抗生成網路建議模式	網路爬蟲對應之內容關鍵字
校園愛情	放學、校門、書包、課本、學長、學姊.....
互看不順眼	殺人、氣、急、不耐、不理不睬、不悅.....
日久生情	發現喜歡上、漸漸喜歡上、陪伴、改觀.....
喜劇	男朋友、女朋友、我也喜歡你、在一起.....
安靜沉穩	靦腆、靜靜陪著、面無表情.....
活潑開朗	燦爛笑容、開朗、活力滿滿、沒心沒肺.....
學習諒解	和好、和解、緩和、不再排斥、體諒.....

圖七：國家網路中心遠端計算回傳建議文本組合

圖八：網路爬蟲內容關鍵字資料庫

伍、討論

書名	故事背景(x1)	初次相識(x2)	交往過程(x3)
回到月亮許諾的那天	校園愛情(x11)	互看不順眼(x21)	日久生情
最後的再見	校園→都市愛情(x12)	互看不順眼	日久生情
載著流星的人	校園→都市愛情	重組家庭的兄妹	日久生情
紙星星	校園愛情	互看不順眼	日久生情
首席老公	都市愛情(x13)	互看不順眼	日久生情
來自天堂的雨	校園愛情	女主角的朋友牽線	日久生情
與你相愛的抉擇	校園愛情	男主角注意女主角	日久生情
老公，別越過界	校園→都市愛情	社團認識	舊情復燃
如果你也聽說	校園愛情	男主角注意女主角	相知相惜
天王助理	都市愛情	男主角注意女主角	日久生情
十二夢	校園愛情	男主角愛慕女主角	日久生情

- 言情小說的主要寫作結構分析：
- 推斷因為**背景與讀者高度相關**，故能成功吸引讀者。
- 其引人入勝的情節在於**男女主角關係的重大轉折**，以及最後以**喜劇收尾**的筆法。

伍、討論

書名	探討議題(x8)	男主角是否為初戀(x9)	女主角是否為初戀(x10)
回到月亮許諾的那天	學習諒解，青少年成長，校園霸凌	是	是
最後的再見	把握當下，青少年成長，生死議題，夢想，穿越重生	是	是
載著流星的人	重組家庭，再婚議題，手足情誼，青少年成長	否	否
紙星星	青少年成長，手足情誼，青少年犯罪，心理疾病，家庭問題，性別議題，夢想	是	是
首席老公	家庭問題，幼兒教育，婚姻議題，夢想	是	是
來自天堂的雨	珍惜生命，校園霸凌，青少年成長	是	否
與你相愛的抉擇	青少年成長，校園霸凌，心理疾病，家庭問題，性別議題，珍惜生命，穿越重生	否	是
老公，別越過界	家庭問題，幼兒教育，婚姻議題	是	是
如果你也聽說	青少年成長，家庭問題，精神虐待，校園霸凌，手足情誼	是	是
天王助理	心理疾病，夢想，網路暴力	是	否
十二夢	青少年成長，校園霸凌，心理疾病，家庭問題，多重人格，校園性騷擾	是	是

伍、討論

書名	男主角特質(x6)	女主角特質(x7)
回到月亮許諾的那天	霸氣(x61), 害羞(x62)	富正義感(x71)
最後的再見	安靜沉穩(x63), 壓抑(x64), 霸氣, 優秀帥氣(x65)	霸氣中性(x72), 堅強倔強(x73), 開朗活潑(x74), 直來直往(x75)
載著流星的人	安靜沈穩, 溫柔可靠(x66), 善於傾聽(x67)	運動健將(x76), 堅強倔強, 直來直往
紙星星	話少神秘(x68), 兇狠(x69), 安靜沉穩, 叛逆(x610), 頭痛人物(x611)	優秀, 懦弱聽話, 功課優秀, 多才多藝, 模範生
首席老公	安靜沉穩, 優秀帥氣, 多金總裁(x612), 溫柔可靠	外貌亮麗, 優秀, 善解人意, 千金小姐, 開朗賢淑, 討人喜愛
來自天堂的雨	安靜沈穩、游泳校隊(x613)、優秀帥氣	活潑開朗, 害怕背叛, 勇於面對
與你相愛的抉擇	安靜沉穩, 溫柔可靠, 優秀帥氣	知性, 個性堅強, 性格善良, 文靜懂事, 平凡膽小
老公, 別越過界	安靜沉穩, 優秀帥氣, 多金總裁	溫柔優雅, 外貌亮麗, 優秀, 善解人意
如果你也聽說	口吃(x614), 膽小懦弱(x615), 細心體貼(x616), 不擅交際(x617)	安靜沉穩, 討人喜愛, 功課優秀, 外貌亮麗, 多才多藝, 模範生
天王助理	開朗活潑(x618), 善良暖男(x619), 優秀帥氣, 偶像歌手(x620)	優雅知性, 外貌亮麗, 功課優秀, 模範生, 情商低
十二夢	膽小懦弱, 畏畏縮縮(x621), 不擅交際	完美, 功課優秀, 性格善良, 外貌亮麗, 模範生

• 吸引人的主要情節結構
成因分析：

1. 推斷出易吸引異性的人格特質。
2. 探討議題越多元時，書籍網路聲量也越高，兩者成正相關。

陸、結論與建議

➤ 結論

• 商業化言情小說情節及劇本設計模式：

1. 參考**時代背景**來進行寫作。
2. 探討**多元化的社會議題**，以期待增加讀者的**認同感**。
3. **網際網路普及**後，增加對**網戀**的描寫。

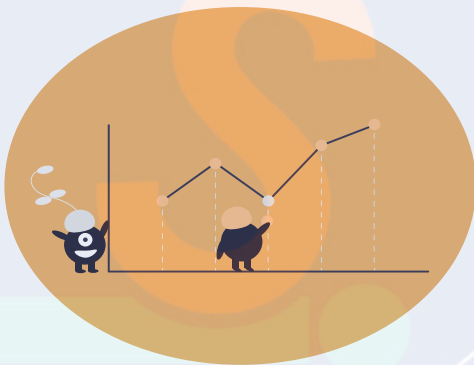
• 以對抗生成網路生成商業化言情小說情節及劇本設計模式之可行性評估：

1. 本研究成功地相關研究目標，具體驗證以GAN生成商業化言情小說情節及劇本設計模式之可行性。

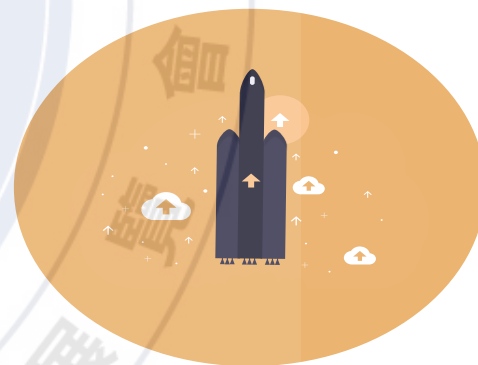
➤ 未來展望



- 以網路爬蟲擴大數據庫，亦即以爬蟲軟體更新不同小說的內容。



- 以文字雲與關鍵字分析方法量測出最佳文本資料。



- 具體提昇非監督學習的對抗生成網路分析計算的精確程度。

參考資料

- [1] 楊若慈。(2015)。那些年，我們愛的步步驚心-台灣言情小說浪潮中的性別政治。台北市：秀威資訊科技股份有限公司。
- [2] 謝奇任。(2016)。致我們的青春：臺灣、日本、韓國與中國大陸的網路小說產業發展。臺北市：威秀經典。
- [3] 倪采青。(2010)。言情小說 編織另類作家夢，Career職場情報誌9月號第413期。
- [4] Martin Halvey and Mark T. Keane(2007), An Assessment of Tag Presentation Techniques, poster presentation at WWW.
- [5] 趙維孝。(2020)如何快速掌握社群議題趨勢 – 從文字雲解析4個月東京奧運聲量！。資料取自：
<https://blog.qsearch.cc/2020/03/tokyo2020/>
- [6] Caesar, Holger. (2019). A list of papers on Generative Adversarial (Neural) Networks: nightrome/really-awesome-gan.