

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號： 97/29796

※申請日期： 97.08.06 ※IPC 分類： G06F 11/29 (2006.01)

一、發明名稱：(中文/英文)

文章產生系統及方法/ARTICLE GENERATING SYSTEM AND METHOD

二、申請人：(共1人)

姓名或名稱：(中文/英文)(簽章) ID : 46804706

國立交通大學/ National Chiao Tung University

指定 為應受送達人

代表人：(中文/英文)(簽章) /

住居所或營業所地址：(中文/英文)

300 新竹市大學路 1001 號/

No.1001, Dasyue Rd., Hsinchu City 300, Taiwan (R.O.C.)

國 籍：(中文/英文) 中華民國/TW

三、發明人：(共3人)

姓 名：(中文/英文)

1. 李嘉晃/ LEE, CHA-HOANG

2. 余思翰/ YU, SZU-HAN

3. 陳智維/ CHEN, CHIH-WEI

國 籍：(中文/英文)

1. 中華民國/TW

2. 中華民國/TW

3. 中華民國/TW

四、聲明事項：

主張專利法第二十二條第二項第一款或第二款規定之事實，其事實發生日期為： 年 月 日。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

五、中文發明摘要：

本發明提供一種文章產生系統。該文章產生系統包含綜合語料庫、特定主題語料庫、判斷模組、決定模組、選取模組及組合模組。當特定主題被決定時，判斷模組根據綜合語料庫與特定主題語料庫產生對應於特定主題之複數個主題關鍵詞。決定模組根據複數個主題關鍵詞從特定主題語料庫中決定複數個填充字串。選取模組自複數個主題關鍵詞中選出複數個目標關鍵詞，並根據複數個目標關鍵詞從複數個填充字串中選出複數個候選填充字串。組合模組根據複數個目標關鍵詞與複數個候選填充字串形成與特定主題相關之文章。

六、英文發明摘要：

The invention provides an article generating system. The article generating system comprises a general word database, a specific topic word database, a judging module, a determining module, a selecting module, and a combining module. When a specific topic is determined, the judging module generates topic keywords corresponding to the specific topic according to the general word database and the specific topic word database. The determining module determines a plurality of filling words from the specific topic word database according to the topic keywords. The selecting module selects target keywords from the topic keywords and selects candidate filling words from the filling words according to the target keywords. The combining module forms an article related to the specific topic according to the target keywords and the candidate filling words.

七、指定代表圖：

(一)本案指定代表圖為：第(一)圖。

(二)本代表圖之元件符號簡單說明：

1：文章產生系統

10：綜合語料庫

12：特定主題語料庫

14：判斷模組

16：決定模組

18：選取模組

20：組合模組

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

九、發明說明：

【發明所屬之技術領域】

本發明係與文章產生有關，並且特別地，本發明係關於一種可自動產生與特定主題相關之文章的文章產生系統及方法。

【先前技術】

近年來，隨著人工智慧領域之技術不斷地發展，其在數位學習上之應用亦相當廣泛，例如：電腦問答系統、電腦自動作文評分系統等。由於文章之寫作乃屬於一項具有高度技巧性之工作，再加上現代人對於寫作效率之要求日益提高，因此，能夠有效地運用人工智慧以輔助使用者進行寫作之文章自動產生系統或寫作輔助系統相關之研究亦逐漸開始受到重視。

一般而言，此類寫作輔助系統大多採用包含龐雜且具有各種不同主題的文件之綜合語料庫作為產生文章之素材來源。然而，此類寫作輔助系統之缺點在於其運用綜合語料庫之素材所產生之文章未必能具有一貫的主題性且文章的結構及內容很可能與這些素材之相似度太高。因此，當使用者欲使用此類系統輔助其產生與一特定主題相關之文章時，實際上所得到之文章卻往往與主題關連性不高或通篇之語句均相當不通順，故無法滿足使用者之需求。

此外，由於目前中文已經成為世界上使用人口最多的語言之一，並且全球各地想要學習中文的人數亦陸續地增加中。然而，目前仍缺乏一套完善且易操作之文章自動產生系統或寫作輔助系統，以有效地提供中文寫作輔助功能，並產生符合主題性且語句通順之中文文章。對於中文初學者而

言，的確不甚方便。

因此，本發明提供一種文章產生系統及方法，以解決上述問題。

【發明內容】

本發明之範疇在於提供一種可自動產生與特定主題相關之文章的文章產生系統及方法。在實際應用中，該文章產生系統及方法能夠根據使用者所指定之特定主題自動產生文章或是協助使用者進行特定主題文章之寫作。

根據本發明之第一具體實施例為一種文章產生系統。該文章產生系統包含綜合語料庫、複數個主題語料庫、判斷模組、決定模組、選取模組及組合模組。其中該綜合語料庫儲存有複數篇一般主題文章；該複數個主題語料庫包含對應於各種不同主題之語料庫，其中之特定主題語料庫儲存有與特定主題相關之複數篇特定主題文章。

於此實施例中，當特定主題被使用者或系統決定時，判斷模組將會根據綜合語料庫與特定主題語料庫產生對應於特定主題之複數個主題關鍵詞。接著，決定模組根據複數個主題關鍵詞從複數篇特定主題文章中決定複數個填充字串；選取模組則會自複數個主題關鍵詞中選出複數個目標關鍵詞，並根據複數個目標關鍵詞從複數個填充字串中選出複數個候選填充字串。最後，組合模組根據複數個目標關鍵詞與複數個候選填充字串形成與特定主題相關之文章。

根據本發明之第二具體實施例為一種文章產生方法。首先，該文章產生方法決定一特定主題並根據綜合語料庫與特定主題語料庫產生對應於特定主題之複數個主題關鍵詞。其中特定主題語料庫儲存有與特定主題相關之複數篇特定主題

文章；綜合語料庫則儲存有複數篇一般主題文章。接著，該文章產生方法根據複數個主題關鍵詞從複數篇特定主題文章中決定複數個填充字串，並自複數個主題關鍵詞中選出複數個目標關鍵詞。然後，該文章產生方法根據複數個目標關鍵詞從複數個填充字串中選出複數個候選填充字串。最後，該文章產生方法根據複數個目標關鍵詞與複數個候選填充字串形成與特定主題相關之文章。

根據本發明之第三具體實施例係一種情書自動產生系統。此情書自動產生系統與第一具體實施例的文章產生系統最大不同之處在於：自複數個主題關鍵詞中選出複數個目標關鍵詞之選取方式及關鍵詞串列之組合與擴展方式。

於此實施例中，當一特定主題文章之複數個主題關鍵字已被選出後，該情書自動產生系統透過一種改良後的 SPLR (strict phrase likelihood ratio) 方法分別計算每一個主題關鍵詞之 SPLR 分數，只有當某一主題關鍵詞之 SPLR 分數高於一臨界值時，該主題關鍵詞才會選為目標關鍵詞。接著，該情書自動產生系統可將該主題文章中之各個目標關鍵詞組合成關鍵詞串列。

此外，為了能夠增加關鍵詞串列之關鍵詞數目，該情書自動產生系統即可分別將該文章中位於每一個目標關鍵詞之前後各 N 個主題關鍵詞亦視為目標關鍵詞，其中 N 與該關鍵詞串列所欲包含之總關鍵詞數目有關，N 亦可由使用者決定。

相較於先前技術，根據本發明之文章產生系統及方法能夠有效地提供使用者中文寫作輔助之功能，並能自動產生符合主題性且語句通順之文章，可幫助使用者透過有效率且易操作之方式學習中文寫作或產生主題性之文章。因此，該文

章產生系統及方法除了能夠大幅地提升中文寫作效率外，還具有廣大中文學習人口之市場潛力。

關於本發明之優點與精神可以藉由以下的發明詳述及所附圖式得到進一步的瞭解。

【實施方式】

本發明提供一種文章產生系統及方法。該文章產生系統可以輔助使用者進行文章之寫作並可自動根據一特定主題產生與該特定主題相關之文章。

根據本發明之第一具體實施例係一種文章產生系統。該文章產生系統之主要運作原理為：先自與特定主題相關之複數篇特定主題文章中擷取出複數個關鍵字詞並將這些關鍵字詞串列成文章之基本骨架。接著，再根據每兩個相鄰之關鍵字詞自複數篇特定主題文章中選出較適合填入該兩個相鄰關鍵字詞之間的複數個候選填充字詞。此時，使用者可以依其喜好或判斷自這些候選填充字詞中挑選出最適合的填充字詞，並將其填入該兩個相鄰關鍵字詞之間。

請參照圖一，圖一係繪示該文章產生系統之功能方塊圖。如圖一所示，文章產生系統 1 包含綜合語料庫 10、特定主題語料庫 12、判斷模組 14、決定模組 16、選取模組 18 及組合模組 20。實際上，文章產生系統 1 除了特定主題語料庫 12 之外，還包含許多個分別對應於不同主題之主題語料庫。至於綜合語料庫 10 則可以是一個大型的平衡語料庫，例如中研院所提供包含約五百萬個字詞的平衡語料庫。

如圖一所示，於文章產生系統 1 中，判斷模組 14 係耦接至綜合語料庫 10 及特定主題語料庫 12；決定模組 16 係耦接至特定主題語料庫 12 及判斷模組 14；選取模組 18 係耦接至

判斷模組 14 及決定模組 16；組合模組 20 係耦接至選取模組 18。接下來，將分別就文章產生系統 1 所包含之各模組及其具備的功能進行詳細之介紹。

於此實施例中，綜合語料庫 10 儲存有複數篇一般主題文章；特定主題語料庫 12 儲存有與一特定主題相關之複數篇特定主題文章。首先，需要決定欲產生之文章所對應的特定主題。實際上，判斷模組 14 可包含主題選擇裝置(未顯示於圖中)以供使用者從複數個候選主題中選出特定主題。此外，特定主題亦可由文章產生系統 1 自動產生。

當特定主題被決定後，判斷模組 14 即會根據綜合語料庫 10 與特定主題語料庫 12 產生對應於該特定主題之複數個主題關鍵詞。接下來，將就判斷模組 14 如何根據綜合語料庫 10 與特定主題語料庫 12 產生複數個主題關鍵詞進行介紹。

首先，假設使用者已決定了特定主題，判斷模組 14 即會根據特定主題自特定主題語料庫 12 所儲存之複數篇特定主題文章中，經過斷詞與詞性標記等處理步驟以整理出與特定主題相關之複數個詞語單元。值得注意的是，所有的標點符號亦會被視為詞語單元。

舉例而言，假設使用者自一選擇視窗所顯示之眾多候選主題中選擇特定主題為「下課十分鐘」，若對應「下課十分鐘」之特定主題語料庫 12 總共儲存有一百篇與「下課十分鐘」相關之文章，則判斷模組 14 將會從這一百篇文章中經過斷詞與詞性標記等處理步驟，整理出複數個詞語單元。很明顯地，由於這一百篇文章係與「下課十分鐘」之主題有關，因此，從這一百篇文章中所得到的複數個詞語單元自然也與「下課十分鐘」之主題有關。

接著，將介紹判斷模組 14 如何從上述這些詞語單元中選出複數個主題關鍵詞。顧名思義，所謂的「主題關鍵詞」應該是在特定主題文章中較常出現且相對地在一般主題文章中較少出現之詞語。因此，判斷模組 14 可以根據某一詞語單元在特定主題文章中與一般主題文章中出現頻率的差異大小來決定該詞語單元是否為主題關鍵詞。也就是說，若某一詞語單元在特定主題文章中出現頻率很高而在一般主題文章中出現頻率很低，該詞語單元即很有可能是判斷模組 14 所要尋找的主題關鍵詞。

在實際應用中，假設特定主題語料庫 12 具有將所有詞語單元根據其於所有特定主題文章中出現頻率之高低進行排列而得之第一排序，而綜合語料庫 10 具有將所有詞語單元根據其於所有一般主題文章中出現頻率之高低進行排列而得之第二排序。若同一詞語單元於第一排序中具有第一排名且於第二排序中具有第二排名，則判斷模組 14 即可根據該詞語單元之第一排名與第二排名的差距是否大於某一臨界值，以判斷該詞語單元是否可被選為主題關鍵詞。

實際上，排名差距之臨界值亦可由使用者設定。當排名差距之臨界值設得較小時，判斷模組 14 所得到之主題關鍵詞將會相當多，很可能造成許多主題關鍵詞並非真正的「關鍵詞」；若排名差距之臨界值設得非常大，判斷模組 14 所得到的主題關鍵詞很可能會非常少甚至沒有。因此，排名差距之臨界值的設定是否適當關係著判斷模組 14 能否順利得到有用的主題關鍵詞。

舉例而言，假設特定主題為「下課十分鐘」，排名差距之臨界值設為 300，且第一排序與第二排序分別如圖二(A)與圖二(B)所示。由圖二(A)與圖二(B)可知，詞語單元「下課」在

第一排序中排名第二，但在第二排序中則排名第 6359，這代表著「下課」一詞在與「下課十分鐘」主題相關之文章中出現之頻率遠較一般主題文章來得高。由於「下課」在第一排序與第二排序之排名差距大於臨界值 300，因此，判斷模組 14 即會將「下課」選為主題關鍵詞。

相反地，如圖二(A)與圖二(B)所示，詞語單元「的」在第一排序與第二排序中均排名第一，這代表「的」一詞無論在與「下課十分鐘」主題相關之文章或一般主題文章中，出現的頻率均非常高。由於「的」在第一排序與第二排序之排名差距遠小於臨界值 300，因此，判斷模組 14 並不會將「的」選為主題關鍵詞。

於此實施例中，當判斷模組 14 已產生對應於特定主題之複數個主題關鍵詞後，決定模組 16 將會根據該複數個主題關鍵詞從特定主題語料庫 12 之複數篇特定主題文章中決定複數個填充字串。

承上例，假設判斷模組 14 已成功選出與「下課十分鐘」主題相關之主題關鍵詞，其總數為三百個。接著，決定模組 16 即會針對每一篇特定主題文章將其中所有出現主題關鍵詞之部分均加以標示，並將每兩個相鄰的關鍵詞間之文字片段標示為填充字串，藉此，決定模組 16 即可擷取所有被標示之填充字串以得到該複數個填充字串。此外，在每一篇特定主題文章中，全部(或部分)被標示的主題關鍵詞將可形成關鍵詞串列。

舉例而言，於「這短短的十分鐘，對每個人來說或多或少都有一定用處。」這段文字中，假設「這」、「多」、「分鐘」、「用處」及「每個人」均屬於主題關鍵詞，若將所有主題關鍵詞均以底線標示即變成「這短短的十分鐘，對每個人

來說或多或少都有一定用處。」。

此時，於這段文字中形成了「短短的十」、「，對」、「來說或」及「或多或少都有一定」等填充字串；至於這段文字所形成之關鍵詞串列可以是由其全部的主題關鍵詞所構成的「這…分鐘…每個人…多…用處」或由其部分的主題關鍵詞所構成的「分鐘…每個人…多」。

於此實施例中，當決定模組 16 已經決定複數個填充字串之後，選取模組 18 將會自複數個主題關鍵詞中選出複數個目標關鍵詞。接著，選取模組 18 將會再根據其所選出之複數個目標關鍵詞從複數個填充字串中選出複數個候選填充字串。

在實際應用中，選取模組 18 可包含關鍵詞選擇裝置(未顯示於圖中)以供使用者從複數個主題關鍵詞中選出複數個目標關鍵詞。於此實施例中，目標關鍵詞之選取可透過選擇關鍵詞串列之方式進行。舉例而言，由於決定模組 16 已經針對每一篇特定主題文章進行處理以得到複數個關鍵詞串列，故使用者可以從這些關鍵詞串列中選出想要的關鍵詞串列以作為欲產生文章之架構。

值得一提的是，文章產生系統 1 最後所產生之文章長度與使用者所選取之目標關鍵詞數目有關。一般而言，若目標關鍵詞之數目愈多，則文章產生系統 1 產生之文章長度通常亦愈長。因此，使用者亦可藉由調整關鍵詞串列所包含之關鍵詞的數目來達到控制欲產生文章長度之功效。

於此實施例中，當選取模組 18 選出包含複數個目標關鍵詞之關鍵詞串列後，選取模組 18 將會根據該關鍵詞串列中之每兩個相鄰目標關鍵詞從複數個填充字串中擷取出對應於該兩個相鄰目標關鍵詞的複數個候選填充字串，亦即每一個候

選填充字串均曾在特定主題文章中出現於該兩個相鄰目標關鍵詞之間。

為了使此一程序能順利執行，在先前決定模組 16 針對每一篇特定主題文章中之每兩個相鄰的關鍵詞間的文字片段標示為填充字串時，即需要順便記錄每一個填充字串所分別對應之兩個相鄰的關鍵詞，亦即記錄每一個填充字串曾經在特定主題文章中出現於哪兩個關鍵詞之間，以作為選取模組 18 選取候選填充字串時之根據。

在實際應用中，為了使用者操作上的方便，文章產生系統 1 亦可進一步對於這些候選填充字串進行排序並可自動將排序位於最前面的候選填充字串填入該兩個相鄰的目標關鍵詞之間。其排序之方式可以根據這些候選填充字串與該兩個相鄰的目標關鍵詞之語意關連程度或在文章中出現之位置等因素決定。若使用者對於文章產生系統 1 自動填入的填充字串感到不滿意，使用者可再從其他候選填充字串中選出其感到滿意的填充字串。此外，若使用者對於所有的候選填充字串均不滿意，使用者還可透過自行輸入之方式輸入適當的填充字串。

藉由上述的方式選出複數個目標關鍵詞以及對應於每兩個相鄰目標關鍵詞的複數個候選填充字串後，組合模組 20 即可根據這些目標關鍵詞與候選填充字串形成與特定主題相關之文章。

接下來，將藉由圖三(A)至圖三(E)等圖示就文章產生系統 1 實際的操作視窗介面及操作流程做一詳細說明。首先，使用者將會進入文章產生系統 1 之文章主題選擇視窗，如圖三(A)所示。於此視窗中，當使用者從不同的文章主題中選出欲產生文章之主題(例如下課十分鐘)並按下 OK 的按鈕後，即

會進入選擇文章長度之視窗，如圖三(B)所示。

在選擇文章長度之視窗中，使用者可選擇欲產生長度較短、中等或較長的文章。由於文章長度與其所包含之關鍵字數目有關，所以不同長度之文章所包含的關鍵字數目亦不相同，例如 10 個、20 個或 30 個關鍵字。於此例中，假設使用者選擇產生較短之文章(包含 10 個關鍵字)並按下 OK 的按鈕後，即會進入到選擇關鍵詞串列之視窗，如圖三(C)所示。

於選擇關鍵詞串列之視窗中，系統將會預設選擇排序第一之關鍵詞串列，但使用者亦可自由地從畫面中所列出之其他各組關鍵詞串列裡進行選擇。當使用者完成選擇關鍵詞串列之動作並按下 OK 的按鈕後，即會進入圖三(D)所示之文章編輯畫面。

如圖三(D)所示，於文章編輯畫面中，系統將會先根據使用者所選擇之關鍵詞串列 1「短短…十…分鐘…寶貴…通常…去…廁所…就是…找…老師」自動產生一篇預設文章「短短的十分鐘，非常的寶貴。通常做的事不是去廁所，就是去找老師，根本無法做其他的事。」並將其顯示於文章編輯視窗內。此外，文章編輯畫面之上方顯示的是預設文章中所包含的 10 個關鍵詞及填充詞標籤 F1~F10。

假設使用者對於預設文章中關鍵詞「分鐘」與「寶貴」之間所填入的「，非常的」感到不甚滿意，使用者即可點選填充詞標籤 F3，文章編輯畫面中之填充詞候選視窗即會列出適合填入填充詞標籤 F3 並已經過排序之複數個候選填充詞以供使用者選擇，如圖三(E)所示。其中排在第一個的候選填充詞即為預設文章中所選用的「，非常的」。此外，若使用者對於所有候選填充詞均不甚滿意，使用者亦可選擇填充詞候選視窗中所顯示之第五個選項<<自行輸入>>。此時，文章編輯

畫面中將會出現一輸入視窗以供使用者自行輸入文字以作為填入填充詞標籤 F3 之填充詞。

在實際應用中，當文章產生系統 1 提供使用者候選填充詞時，亦可進一步增加同義詞之功能。舉例而言，若原本預設文章所選用的填充詞為「可能」，假設使用者想要使用「可能」之同義詞，使用者即可選擇同義詞之功能。此時，文章產生系統 1 將會列出「可能」之各個候選同義詞，例如「或許」、「大概」、「也許」等，使用者即可在這些候選同義詞中進行選擇。

此外，文章產生系統 1 亦可提供譬喻之修辭功能。舉例而言，若候選填充詞為一喻詞「好像」，表示此處可能運用到譬喻中之明喻的修辭技巧，因此，文章產生系統 1 亦可列出與「好像」類似的其他喻詞，例如「彷彿」、「似乎」、「就像」等，以供使用者選擇。

綜上所述，文章產生系統 1 可以自動根據使用者所指定之文章主題及長度(關鍵詞數目)產生與該主題相關之文章，而使用者亦可針對文章的內容進行調整以使文章能更符合使用者之需求，因此，文章產生系統 1 可達到輔助使用者進行文章寫作之具體功效。

根據本發明之第二具體實施例係一種文章產生方法。該文章產生方法能夠輔助使用者進行文章之寫作並可自動產生與特定主題相關之文章。請參照圖四，圖四係繪示該文章產生方法之流程圖。

如圖四所示，首先，該方法執行步驟 S10，決定特定主題。實際上，特定主題可由使用者從複數個候選主題中選出。接著，該方法執行步驟 S12，根據綜合語料庫與特定主

題語料庫產生對應於該特定主題之複數個主題關鍵詞。其中特定主題語料庫儲存有與特定主題相關之複數篇特定主題文章，綜合語料庫可以是大型的平衡語料庫並儲存有複數篇一般主題文章。

於實際應用中，該方法可以根據某一詞語單元在特定主題文章中與一般主題文章中出現頻率之差異大小來決定該詞語單元是否為主題關鍵詞。也就是說，若某一詞語單元在特定主題文章中出現頻率很高而在一般主題文章中出現頻率很低，即代表該詞語單元很有可能即為主題關鍵詞。

於此實施例中，假設特定主題語料庫具有將所有詞語單元根據其於所有特定主題文章中出現頻率之高低進行排列而得之第一排序，而綜合語料庫具有將詞語單元根據其於所有一般主題文章中出現頻率之高低進行排列而得之第二排序。若同一詞語單元於第一排序中具有第一排名且於第二排序中具有第二排名，該方法即可根據該詞語單元之第一排名與第二排名的差距是否大於某一臨界值，以判斷該詞語單元是否可被選擇為主題關鍵詞。

舉例而言，假設排名差距之臨界值為 400，若詞語單元「機械」在第一排序中之第一排名為 2 且在第二排序中之第二排名為 2434，由於第一排名與第二排名的差距很明顯大於臨界值，因此，詞語單元「機械」將會被選為主題關鍵詞。

當複數個主題關鍵詞已被選出後，該方法將會執行步驟 S14，根據該複數個主題關鍵詞從複數篇特定主題文章中決定複數個填充字串。如圖五所示，於步驟 S14 中，該方法可先執行子步驟 S141，針對每一篇特定主題文章將所有出現關鍵詞之部分均標示起來。接著，該方法再執行步驟 S142，將每兩個相鄰的關鍵詞間之文字片段均標示為填充字串，藉此，

該方法即可擷取所有被標示之填充字串以得到該複數個填充字串。

接下來，該方法執行步驟 S16，自該複數個主題關鍵詞中選出複數個目標關鍵詞。實際上，該複數個目標關鍵詞可以由使用者從複數個主題關鍵詞中選出。接著，該方法執行步驟 S18，根據步驟 S16 所選出的複數個目標關鍵詞，從該複數個填充字串中選出複數個候選填充字串。

最後，該方法執行步驟 S20，根據該複數個目標關鍵詞與該複數個候選填充字串形成與該特定主題相關之文章。在實際應用中，使用者可根據該複數個目標關鍵詞選出複數個替代目標關鍵詞，並根據該複數個替代目標關鍵詞與該複數個候選填充字串形成與特定主題相關之文章。此外，使用者亦可根據該複數個候選填充字串選出複數個替代填充字串，並根據複數個目標關鍵詞與該複數個替代填充字串形成與特定主題相關之文章。

根據本發明之第三具體實施例係一種情書自動產生系統。值得注意的是，此情書自動產生系統與第一具體實施例的文章產生系統最大不同之處在於：自複數個主題關鍵詞中選出複數個目標關鍵詞之選取方式及關鍵詞串列之組合與擴展方式。接下來，將分別就這兩點進行探討。

首先，關於該情書自動產生系統自複數個主題關鍵詞中選取目標關鍵詞之方式，主要係透過一種改良後的 SPLR 方法分別計算每一個主題關鍵詞之 SPLR 分數，只有當某一主題關鍵詞之 SPLR 分數高於一臨界值時，該主題關鍵詞才會選為目標關鍵詞。

所謂的 SPLR 方法一般係應用於找出未知字(unknown

word)之技術上，為了將 SPLR 方法應用於情書自動產生系統，經過多次實驗後，發現到若將原本 SPLR 中之遞迴式的方法去除，所得到的效果最佳。因此，情書自動產生系統即利用改良後的 SPLR 方法，於複數個主題關鍵詞中找出真正重要且能代表此篇文章的目標關鍵詞，再將這些目標關鍵詞展現出來以供使用者選取。改良後的 SPLR 方法可以用下列公式表示：

$$SPLR = \frac{tf(KW_i)}{\text{Max}(tf(KW_{i_L}), tf(KW_{i_R}))}, KW_{i_len} > 1 \quad (1)$$

$$KW_{i_L} = \left\{ KW_{i_L} \left| \left\lceil \frac{KW_{i_len}}{2} \right\rceil \right. \right\} \quad (2)$$

$$KW_{i_R} = (KW_i - \{ KW_{i_L} \}) \quad (3)$$

其中 KW_i 代表關鍵詞； KW_{i_len} 代表關鍵詞字數； KW_{i_L} 代表擷取關鍵詞之左半邊； KW_{i_R} 代表擷取關鍵詞之右半邊； tf 代表在語料庫中出現的頻率。

因此，該情書自動產生系統即是利用上述公式(1)至(3)來計算每一個主題關鍵詞的 SPLR 分數。然而，有時候會遇到一種特殊情況，亦即當該篇文章經過 SPLR 方法進行評分後，並沒有任何主題關鍵詞得到 1 分。因此，該情書自動產生系統即會將所有主題關鍵詞的 SPLR 分數作正規化(normalizing)之處理，以使得所有主題關鍵詞的 SPLR 分數都會分佈於 0 至 1 之間。

於此實施例中，經過上述的 SPLR 評分後，每一個主題關鍵詞將會得到一個介於 0 與 1 間之分數。其中，由於少於兩個字的主題關鍵詞能夠讓使用者感受之意境並不高，故所有少於兩個字的主題關鍵詞將會得到 0 分。之後，該情書自

動產生系統即會根據每一個主題關鍵詞之 SPLR 分數擷取其中得分最高的五個主題關鍵詞作為目標關鍵詞。

此外，經過多次的實驗後，發現到目標關鍵詞的 SPLR 分數臨界值可以設為 0.3 分。也就是說，唯有當某一主題關鍵詞之 SPLR 分數大於 0.3 分時，該主題關鍵詞才會被擷取為情書文章中主要架構上的目標關鍵詞；反之，若主題關鍵詞之 SPLR 分數小於 0.3 分，該主題關鍵詞即不會被擷取為目標關鍵詞。

接下來，將以在某一篇主題文章中所選出之兩個主題關鍵詞「酸甜苦辣」及「學生會長」做為實際的例子來說明上述的 SPLR 評分方式。

首先，在主題關鍵詞「酸甜苦辣」這例子中，若只擷取主題關鍵詞之左半邊，可得到「酸甜苦」；若只擷取主題關鍵詞之右半邊，則可得到「甜苦辣」。假設於該篇文章中，「酸甜苦辣」、「酸甜苦」及「甜苦辣」均出現 10 次，則主題關鍵詞「酸甜苦辣」所得到之 SPLR 分數為 $10/10 = 1$ 分。

另外，在主題關鍵詞「學生會長」這例子中，若只擷取主題關鍵詞之左半邊，可得到「學生會」；若只擷取主題關鍵詞之右半邊，則可得到「生會長」。假設於該篇文章中，「學生會長」共出現 20 次，「學生會」共出現 40 次，「生會長」則出現 20 次，則主題關鍵詞「學生會長」所得到之 SPLR 分數即為 $20/40 = 0.5$ 分。若以上述兩例子作為比較，主題關鍵詞「酸甜苦辣」所得到之 SPLR 分數較「學生會長」來得高，假設 SPLR 分數臨界值為 0.3 分，則兩者均可被視為目標關鍵詞。然而，若 SPLR 分數臨界值被設為 0.6 分，則「酸甜苦辣」可被視為目標關鍵詞，但「學生會長」則否。

於此實施例中，該情書自動產生系統可透過上述 SPLR 評分方法，於每篇文章所包含的眾多主題關鍵詞中選出複數個目標關鍵詞，並可將這些目標關鍵詞形成關鍵詞串列。舉例而言，該情書自動產生系統可以隨機產生 10 個關鍵詞串列以供使用者選取。其主要功能是希望能讓使用者能直接從關鍵詞串列之各目標關鍵詞中體會其意境，不用看過多的資訊，就可以決定自己想產生的文章內容。

接著，將介紹該情書自動產生系統擴展關鍵詞串列之方法。舉例而言，假設透過上述 SPLR 評分方法所得到之關鍵詞串列共包含 5 個目標關鍵詞。為了能夠增加關鍵詞串列之關鍵詞數目，該情書自動產生系統即可分別將該文章中位於每一個目標關鍵詞之前後各 N 個主題關鍵詞亦視為目標關鍵詞，其中 N 與該關鍵詞串列所欲包含之總關鍵詞數目有關， N 亦可由使用者決定。

值得注意的是，該情書自動產生系統係採取交集之方式以避免當不同的目標關鍵詞在取其前後各 N 個主題關鍵詞時，剛好取到相同的主題關鍵詞，因而造成重複選取之現象發生。

接下來，將藉由圖六(A)及圖六(B)就情書自動產生系統實際的操作視窗介面及操作流程做一詳細說明。首先，如圖六(A)所示，使用者可以透過此視窗選取所欲產生之情書內容的長度、所需之關鍵詞串列並輸入寄信者及收信者之名稱。當使用者按下 OK 後，即會進入圖六(B)之視窗。

如圖六(B)所示，情書自動產生系統根據使用者所選取之關鍵詞串列自動產生一篇情書。若使用者不滿意系統自動產生的情書內容，除了可以按下隨機產生情書鍵以重新產生情書外，使用者亦可選擇重選關鍵詞、自創新詞、重選/替換/刪

除某字詞或同義修辭擴展等功能，以將原本情書修改成符合使用者期待之內容。

相較於先前技術，根據本發明之文章產生系統及方法能夠有效地提供使用者中文寫作輔助之功能，並能自動產生符合主題性且語句通順之中文文章，可幫助使用者透過有效率且易操作之方式學習中文寫作或產生主題文章。因此，該文章產生系統除了能夠大幅地提升中文寫作效率外，亦十分具有市場潛力。

藉由以上較佳具體實施例之詳述，係希望能更加清楚描述本發明之特徵與精神，而並非以上述所揭露的較佳具體實施例來對本發明之範疇加以限制。相反地，其目的是希望能涵蓋各種改變及具相等性的安排於本發明所欲申請之專利範圍的範疇內。因此，本發明所申請之專利範圍的範疇應該根據上述的說明作最寬廣的解釋，以致使其涵蓋所有可能的改變以及具相等性的安排。

【圖式簡單說明】

圖一係繪示根據本發明之第一具體實施例的文章產生系統之功能方塊圖。

圖二(A)及圖二(B)分別為一範例中之第一排序與第二排序。

圖三(A)至圖三(E)係分別繪示於一範例中之各操作視窗介面。

圖四係繪示根據本發明之第二具體實施例的文章產生方法之流程圖。

圖五係繪示圖四中之步驟 S14 之詳細流程圖。

圖六(A)及圖六(B)係分別繪示根據本發明之第三具體實施例之情書自動產生系統的各操作視窗介面。

【主要元件符號說明】

S10 ~ S142：流程步驟

1：文章產生系統

10：綜合語料庫

12：特定主題語料庫

14：判斷模組

16：決定模組

18：選取模組

20：組合模組

F1 ~ F10：填充詞標籤

十、申請專利範圍：

- 1、一種文章產生方法，包含下列步驟：
 - (a)決定一特定主題；
 - (b)根據一綜合語料庫與一特定主題語料庫產生對應於該特定主題之複數個主題關鍵詞，其中該特定主題語料庫儲存有與該特定主題相關之複數篇特定主題文章，該綜合語料庫儲存有複數篇一般主題文章；
 - (c)根據該複數個主題關鍵詞，從該複數篇特定主題文章中決定複數個填充字串；
 - (d)自該複數個主題關鍵詞中選出複數個目標關鍵詞；
 - (e)根據該複數個目標關鍵詞，從該複數個填充字串中選出複數個候選填充字串；以及
 - (f)根據該複數個目標關鍵詞與該複數個候選填充字串，以形成與該特定主題相關之文章。
- 2、如申請專利範圍第1項所述之文章產生方法，其中該綜合語料庫為一大型的平衡語料庫。
- 3、如申請專利範圍第1項所述之文章產生方法，其中於步驟(a)中，該特定主題係由一使用者從複數個候選主題中選出。
- 4、如申請專利範圍第1項所述之文章產生方法，該複數個主題關鍵詞中一第一主題關鍵詞出現於該特定主題語料庫之次數為一第一頻率，該第一主題關鍵詞出現於該綜合語料庫之次數為一第二頻率，其中該第一頻率與該第二頻率之差值大於一臨界值。
- 5、如申請專利範圍第1項所述之文章產生方法，其中步驟(c)包含下列子步驟：
 - (c1)於該複數篇特定主題文章中標示該複數個關鍵詞；以及
 - (c2)將每兩個相鄰的關鍵詞間之一文字片段標示為一填充字串，以得到該複數個填充字串。

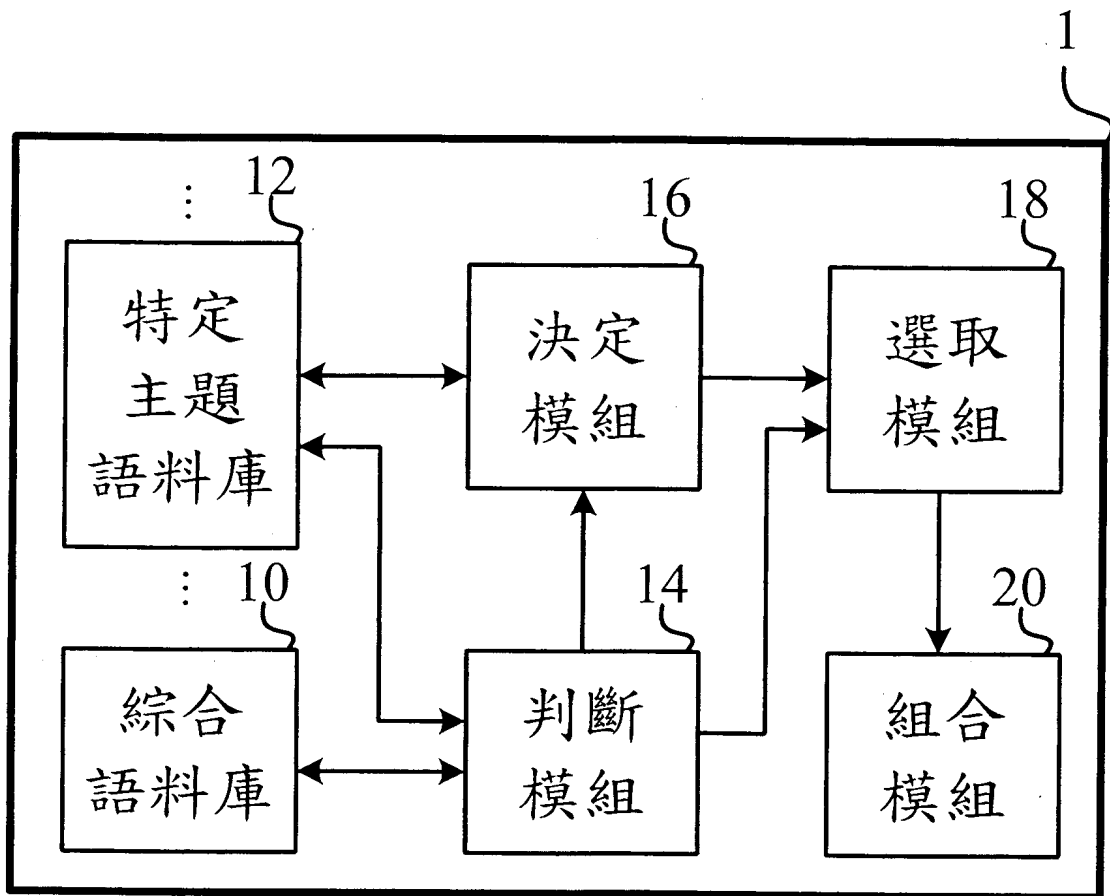
- 6、如申請專利範圍第1項所述之文章產生方法，其中於步驟(d)中，該複數個目標關鍵詞係由一使用者從該複數個主題關鍵詞中選出。
- 7、如申請專利範圍第1項所述之文章產生方法，其中於步驟(d)中，該複數個目標關鍵詞係透過一改良式SPLR(strict phrase likelihood ratio)評分法從該複數個主題關鍵詞中選出。
- 8、如申請專利範圍第1項所述之文章產生方法，其中於步驟(f)中，一使用者根據該複數個目標關鍵詞選出複數個替代目標關鍵詞，並根據該複數個替代目標關鍵詞與該複數個候選填充字串，以形成與該特定主題相關之文章。
- 9、如申請專利範圍第1項所述之文章產生方法，其中於步驟(f)中，一使用者根據該複數個候選填充字串選出複數個替代填充字串，並根據複數個目標關鍵詞與該複數個替代填充字串，以形成與該特定主題相關之文章。
- 10、一種文章產生系統，包含：
 - 一綜合語料庫，該綜合語料庫儲存有複數篇一般主題文章；
 - 複數個主題語料庫，該複數個主題語料庫中之一特定主題語料庫儲存有複數篇特定主題文章，其中該複數篇特定主題文章係與一特定主題有關；
 - 一判斷模組耦接至該綜合語料庫及該複數個主題語料庫，當該特定主題被決定時，該判斷模組根據該綜合語料庫與該特定主題語料庫產生對應於該特定主題之複數個主題關鍵詞；
 - 一決定模組耦接至該複數個主題語料庫及該判斷模組，該決定模組根據該複數個主題關鍵詞從該複數篇特定主題文章中決定複數個填充字串；
 - 一選取模組耦接至該決定模組及該判斷模組，該選取模組自該複數個主題關鍵詞中選出複數個目標關鍵詞，

並根據該複數個目標關鍵詞從該複數個填充字串中選出複數個候選填充字串；以及

一組合模組耦接至該選取模組，該組合模組根據該複數個目標關鍵詞與該複數個候選填充字串形成與該特定主題相關之文章。

- 11、如申請專利範圍第10項所述之文章產生系統，其中該綜合語料庫為一大型的平衡語料庫。
- 12、如申請專利範圍第10項所述之文章產生系統，其中該判斷模組包含一主題選擇裝置產生一選擇介面以供一使用者從複數個候選主題中選出該特定主題。
- 13、如申請專利範圍第10項所述之文章產生系統，其中該判斷模組自該複數篇特定主題文章中整理出複數個詞語單元，若該複數個詞語單元中之一第一詞語單元出現於該特定主題語料庫之次數為一第一頻率，該第一詞語單元出現於該綜合語料庫之次數為一第二頻率，該判斷模組根據該第一頻率與該第二頻率之差值是否大於一臨界值選擇性地決定該第一詞語單元為一主題關鍵詞。
- 14、如申請專利範圍第10項所述之文章產生系統，其中該決定模組於該複數篇特定主題文章中標示該複數個關鍵詞，並將每兩個相鄰的關鍵詞間之一文字片段標示為一填充字串，以得到該複數個填充字串。
- 15、如申請專利範圍第10項所述之文章產生系統，其中該選取模組包含一關鍵詞選擇裝置以供一使用者從該複數個主題關鍵詞中選出該複數個目標關鍵詞。
- 16、如申請專利範圍第10項所述之文章產生系統，其中該選取模組係透過一改良式SPLR評分法從該複數個主題關鍵詞中選出該複數個目標關鍵詞。

十一、圖式：



圖一

排名	字詞
1	的
2	是
3	一
4	在
5	、
6	有
7	個
8	我
9	不
10	了

圖二 (A)

排名	字詞
1	的
2	下課
3	在
4	，
5	有
6	是
7	人
8	分鐘
9	十
10	一

圖二 (B)

請選擇文章主題：

- ◎ 1. 下課十分鐘
- 2. 我的志願
- 3. 我的家庭
- 4. 難忘的旅遊經驗

OK

圖三 (A)

請選擇文章長度：

- ◎ 1. 長度較短的文章(10個關鍵字)
- 2. 長度中等的文章(20個關鍵字)
- 3. 長度較長的文章(30個關鍵字)

OK

圖三 (B)

請選擇關鍵詞串列：

- ◎1. 短短 十 分鐘 寶貴 通常 去 廁所
就是 找 老師
- 2. 學校 下課 可是 熱鬧 鐘 看到 學生
教室 校園 人
- 3. 下課 幾乎 班 玩 同學 些 打球 些
去 廁所
- 4. 下課 學生們 喜歡 課 吧 下課 喜歡 或
是 休息

OK

圖三 (C)

關鍵詞串列：

短短	F1	十	F2	分鐘	F3
寶貴	F4	通常	F5	去	F6
廁所	F7	就是	F8	找	F9
老師	F10				

短短的十分鐘，非常的寶貴。通常做的事不是去廁所，就是去找老師，根本無法做其他的事。

OK

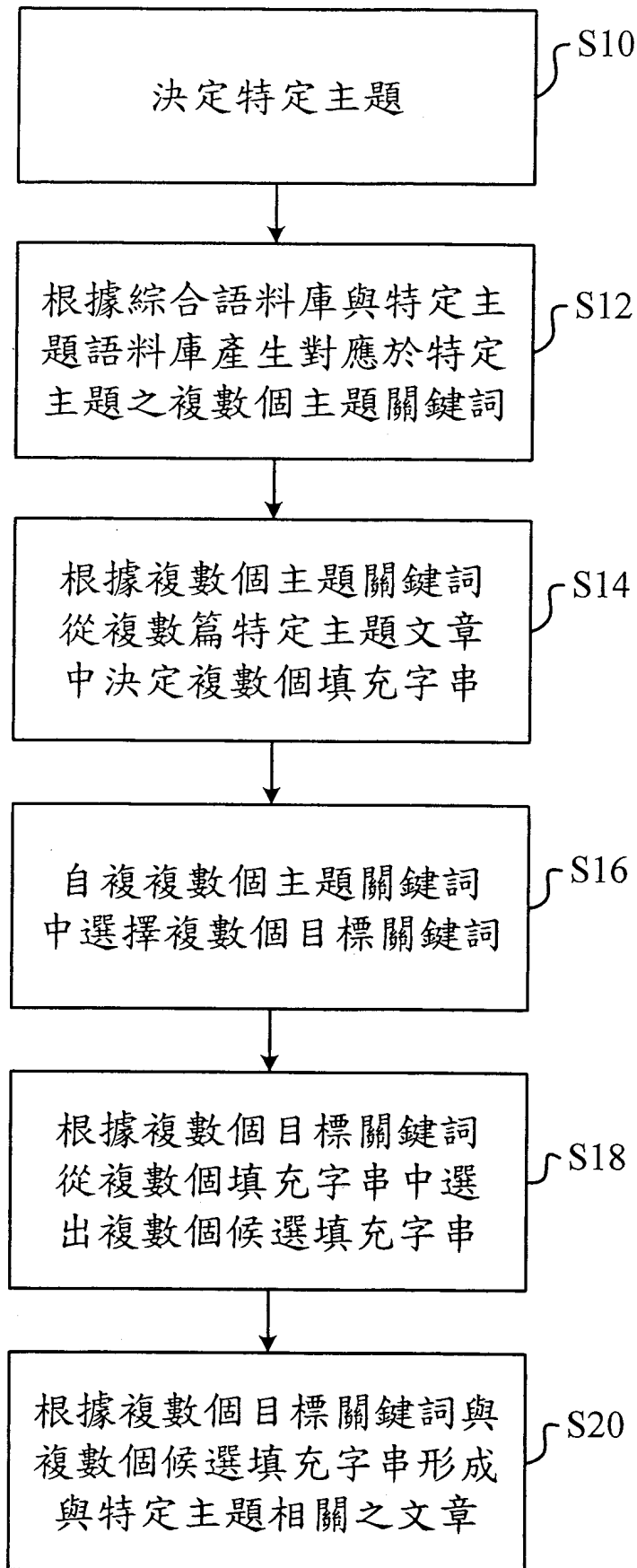
圖三 (D)

F3 之候選填充詞：

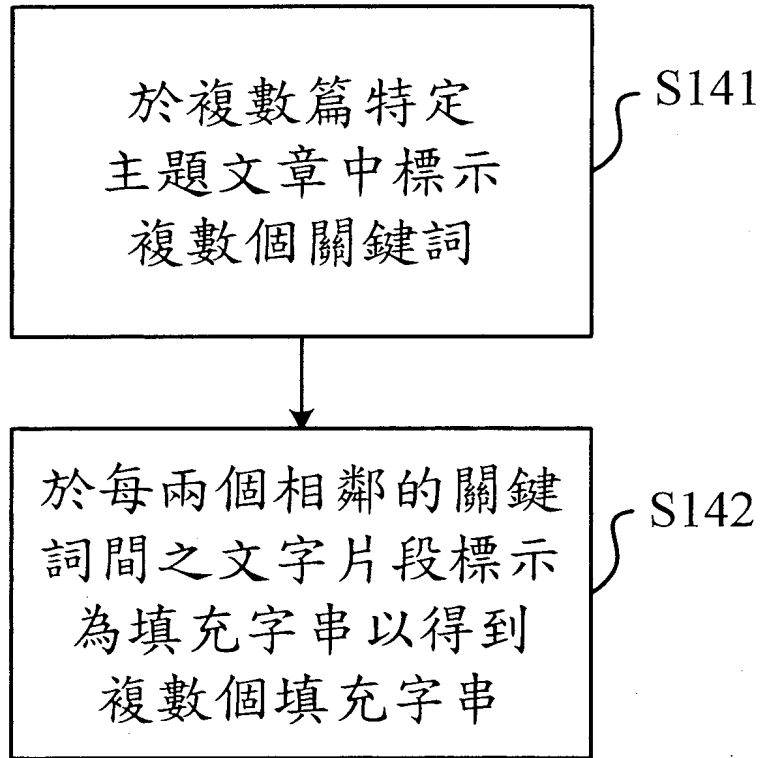
- 1. ，非常的
- 2. 裡，是最
- 3. ，並不能夠做些什麼，但卻很
- 4. 也是很
- 5. <<自行輸入>>

OK

圖三 (E)



圖四



圖五

請選擇關鍵詞串列：

- ◎ 1. 值不值得 葡萄 不謀而合 榆葉梅開 閃閃發光
- 2. 財富 過眼雲煙 籠罩 漫漫 在一起
- 3. 地老天荒 值不值得 執迷不悟 滔滔不絕 臭味相投
- 4. 伴明窗 白頭到老 水汪汪 紅彤彤 一輩子

請選擇情書長度：

- ◎ 1. 短篇情書
- 2. 長篇情書

OK

請輸入寄信者與收信者之名稱：

收信者：

寄信者：

圖六 (A)

情書內容：

親愛的小茹：

我要勇敢的和你在一起，我想我沒有權力害怕失敗，那就請你深入瞭解我後，再看看值不值得你愛……你，就像是一枚青澀的梅子，在江南的雨季裡散發著悠悠的，乾，你是我心裡的葡萄乾。你是我的腸你是我的胃，你永存我心中。

愛妳的小明

候選詞替換：

- 勇敢的
- ◎ 果斷的
- 無論如何
- 馬上
- 立即

重選字詞

替換字詞

自創新詞 刪除字詞

同義修辭擴展 離開程式

重選關鍵詞 隨機產生情書

圖六 (B)

97-12-1

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※ 申請案號：97129796

※ 申請日期：

※IPC 分類：G06F 1/27 (2006.01)

一、發明名稱：(中文/英文)

文章產生系統及方法/ ARTICLE GENERATING SYSTEM AND METHOD

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)(簽章) ID : 46804706

國立交通大學/ National Chiao Tung University

指定 為應受送達人

代表人：(中文/英文)(簽章)/吳重雨/ Wu, Chung-Yu

住居所或營業所地址：(中文/英文)

300 新竹市大學路 1001 號/

No.1001, Dasyue Rd., Hsinchu City 300, Taiwan (R.O.C.)

國 籍：(中文/英文) 中華民國/TW

三、發明人：(共 3 人)

姓 名：(中文/英文)

1. 李嘉晃/ LEE, CHA-HOANG

2. 余思翰/ YU, SZU-HAN

3. 陳智維/ CHEN, CHIH-WEI

國 籍：(中文/英文)

1. 中華民國/TW

2. 中華民國/TW

3. 中華民國/TW