

本功能使用Google Translate翻譯軟體，其翻譯結果僅供參考，本局不保證其翻譯結果之準確、完整或適用於特定目的。

This translation is machine-generated by using Google Translate. The translation is for reference
TIPO does not guarantee the translation is accurate, complete, reliable or fit for specific purposes.

DTD版本：1.0.0

發明專利說明書

※申請案號：095128925

※IPC分類：

一、發明名稱：

字串鏈結方式

A METHOD FOR STRING LINKING

二、中文發明摘要：

本發明之字串鏈結方法，包括：輸入欲進行分類之判別字串；根據所輸入之判別字串，計算各判別字串出現次數；使用一簡稱對照表，將判別字串中特定之簡寫還原；運算兩兩字串間之相似度；將兩字串相似度值運算結果高於一臨界值之兩字串判定為同一字串；以及持續進行相似度比對直至兩兩字串間之相似度均低於所設定之臨界值。

三、英文發明摘要：

A method for string linking comprises inputting determined strings for classifying;calculating the number of each string;restoring contracted form in each string based on a contracted form table;calculating the similarity degree between any two strings;determining the two strings same when similarity degree is over a special degree between the two strings;and comparing the similarity degree between two strings until all similarity degrees are less than the special degree.

四、指定代表圖：

(一)本案指定代表圖為：第1圖

(二)本代表圖之元件符號簡單說明：

101~110 . . . 步驟

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

六、發明說明：

【發明所屬之技術領域】

[0001] 本發明係有關於一種字串鏈結方法，更特別是有關於一種具容錯功能之字串鏈結方法。

【先前技術】

[0002] 一般在進行專利之檢索與評估過程中，當搜尋到一定數量之專利文件後，常需進行專利文件之分類，例如，根據專利之專利權人或是根據專利之名稱進行分類。然，由於名稱可縮寫之緣故，常造成原本應屬相同專利權人之專利文件被分成兩類。

[0003] 例如，以台灣積體電路公司為例，其所登記之專利權人英文名稱，可有TSMC, Taiwan Semiconductor company, Taiwan Semiconductor limited company等均代表台灣積體電路公司，且其中company亦可使用簡寫，CO.，來加以替代。在如此多之登記名稱下，傳統之自動分類方法，會將使用不同專利權人名稱進行登記之專利，即使此些專利本應分類於同一公司名稱，卻將其誤判斷為不同公司所有。因此，傳統之分類方法，在分類完成後，還需藉由使用者之幫助再行將代表同一專利權人，卻被歸屬於不同類之專利，重新分在一起，如此反增加使用者之負擔。

[0004] 因此，如何在增加使用者操作負擔下，增進分類之正確性及成為追求之目標。

【發明內容】

[0005] 因此，本發明之主要目的就是在提供一種字串鏈結方法，可鏈結代表同一目標名稱之各名稱。

[0006] 根據本發明之字串鏈結方法，包括：輸入欲進行分類之判別字串；根據所輸入之判別字串，計算各判別字串出現次數；使用一簡稱對照表，將判別字串中特定之簡寫還原；運算兩兩字串間之相似度；將兩字串相似度值運算結果高於一臨界值之兩字串判定為同一字串；以及持續進行相似度比對直至兩兩字串間之相似度均低於臨界值。

[0007] 根據一實施例，其中具有較少判別字串出現次數之對應字串會被消滅，而存留具有較多判別字串出現次數之對應字串。

[0008] 根據一實施例，運算兩兩字串間之相似度，更包括使用動態規劃演算法，來找出兩字串間之最長共通字串；以及將此共通字串長度進行正規化以獲至一相似度值。

[0009] 綜合上述所言，本發明之方法，可將代表同一事物之各種不同判別名稱進行鏈結並將其歸類入一名稱下，因此，當用於分類時，可將原歸屬不同判別名稱之物併入一名稱下，增進分類之精確性。

【實施方式】

- [0010] 以下將以搜尋並分類以台灣積體電路公司為專利權人專利案之實施例來說明本發明之應用。然值得注意的是，本發明並非僅限於應用在上述之實施例中。
- [0011] 參閱第1圖所示為本發明將不同名稱進行鏈結並歸類為同一名稱之概略流程圖。根據本實施例，台灣積體電路公司其所可能登記之專利權人英文名稱，有TSMC, Taiwan Semiconductor company, Taiwan Semiconductor limited company, Taiwan Semiconductor co., Taiwan Semiconductor LTD. company, 或Taiwan Semiconductor LTD. Co, 等均代表台灣積體電路公司。在傳統分類方法下，這些不同名稱會被判定成不同之專利權人，而使得專利分類失真，而本發明即是用以解決此問題。
- [0012] 首先在步驟101，使用者需輸入欲進行搜尋分類之專利權人名稱，亦即輸入欲進行分類之判別字串。例如，在搜尋出之一定數量專利中，專利分析人員可能想要知道某些專利專利權人名下所擁有之專利，此時即可將該些專利權人之可能名稱輸入。
- [0013] 接著在步驟102中，本發明之方法，即可根據所輸入之判別字串，計算各判別字串出現次數，由於一篇專利專利專利權人名稱僅出現一次，因此計算各專利專利權人名稱次數，亦即計算對應各專利專利權人名稱之專利數目，並將各專利分類至對應之專利權人名稱下。以上述台灣積體電路公司專利權人名稱為例，在此步驟中，會分別計算台灣積體電路公司不同專利權人名稱，例如，TSMC, Taiwan Semiconductor company等專利權人名稱出現次數，並將專利分類至對應之專利權人名稱下。
- [0014] 接著於步驟103，使用一簡稱對照表104，將判別字串中特定之簡寫還原。以上述實施例為例，例如，limited之簡寫為LTD.、company之簡寫為Co. 或Taiwan Semiconductor company之簡寫為TSMC等，這些特定之簡寫與所對應之還原字串，均可由使用者於簡稱對照表104中加以定義，並於步驟103中根據這些簡寫與還原字串間之定義，將判別字串中之簡寫對應還原。例如，Taiwan Semiconductor LTD. Co. 在經由步驟103之還原過程後，其所呈現之字串將成為Taiwan Semiconductor limited company。此時Taiwan Semiconductor LTD. Company和Taiwan Semiconductor LTD. Co在經此還原後，會被視為相同之專利權人。
- [0015] 由於一般在輸入字串時無可避免的會因人為之疏失而造成輸入字串之錯誤，同樣的，在專利申請之過程中，亦有可能因人為之誤繕，而造成專利權人名稱遺漏掉某一字母，例如在輸入「semiconductor」字時，誤輸入成「semiconducter」，而造成應歸屬同一專利專利權人之兩專利，被判定成非屬同一專利專利權人。因此在本發明之步驟105中，會藉由運算兩兩字串間之相似度，並將相似度大於一特定值以上之兩字串視為同一字串，來彌補此些可能發生之人為錯誤。
- [0016] 在一實施例中，本發明判別兩兩字串間相似度之方法，係於步驟107中，藉由使用一習知之動態規劃演算法，來找出兩字串間之最長共通字串，例如，以「Taiwan Semiconductor company」與「Taiwan Semiconducter company」兩字串而言，其中兩字串長度均為26，而兩字串間之最長共通字串為「Taiwan semiconductr company」，其字串長度為25。接著，再於步驟108中，將此共通字串長度進行正規化，例如，將此共通字串長度，25，除以兩字串長度之平均值，26，以獲至一比值96%，亦即相似度值為96%。其中，在經過步驟103還原後之各判別字串，兩兩字串間均會進行上述字串間之相似度運算。
- [0017] 接著，於步驟109，會將兩字串相似度值運算結果高於一臨界值之兩字串判定為同一字串。假設於一實施例中，所設定之臨界值為80%，以上述之例子而言其相似度值為96%，大於所設定之臨界值為80%。因此，「Taiwan Semiconductor company」和「Taiwan Semiconducter company」兩字串會被視為同一字串，此時，分別對應上述字串之兩組專利會彼此合併，而成為一組專利。根據本實施例，由於因誤繕所造成之專利權人名稱錯誤，分類在此誤繕之專利權人名稱下之專利數目畢竟較少。因此，於步驟109中之兩字串對應專利彼此合併，其中具有較少專利數目之對應字串會被消滅，而存留具有較多專利數目之對應字串，易言之，具較少專利數目之專利組會併入具較多專利數目之專利組內。上述之相似度比對與對應專利之合併會持續進行，直至兩兩字串間之相似度均低於所述定之臨界值，如步驟110所述。
- [0018] 值得注意的是，在本發明之相似度比對中，兩字串，如第一與第二字串間，雖然彼此間之相似度值係小於臨界值，但可透過另一字串，如第三字串，因第一與第三字串間，以及第二與第三字串間，彼此之相似度值均大於臨界值，而使得第一與第二字串被視為同一字串，其對應之專利得以合併一起，而視為同一專利權人所有。
- [0019] 綜合上述所言，本發明之方法，可將代表同一事物之各種不同判別名稱進行鏈結並將其歸類入一名稱下，因此，當用於專利分類時，可將原歸屬不同判別名稱下之專利合併入一名稱下，增進專利分類之精確性。
- [0020] 雖然本發明已以一較佳實施例揭露如上，然其並非用以限定本發明，任何熟習此技藝者，在不脫離本發明之精神和範圍內，當可作各種之更動與潤飾，因此本發明之保護範圍當視後附之申請專利範圍所界定者為準。

【圖式簡單說明】

- [0022] 為讓本發明之上述和其他目的、特徵、優點與實施例能更明顯易懂，所附圖式之詳細說明如下：第1圖所示為本發明將不同名稱進行鏈結並歸類為同一名稱之概略流程圖。

【主要元件符號說明】

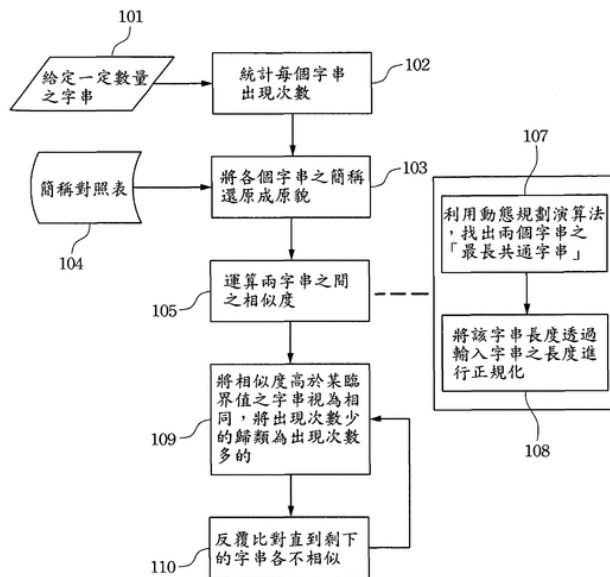
- [0021] 101~110... 步驟

七、申請專利範圍：

1. 一種字串鏈結方法，係用於判別複數個不同字串中應歸屬為同一字串之方法，包括：使用一簡稱對照表，將存於該些字串中特定之簡寫還原；運算兩兩字串間之相似度值；將兩字串間相似度值高於一臨界值之兩字串判定為同一字串；以及持續進行相似度值比對直至任兩字串間之相似度均低於該臨界值。
2. 如申請專利範圍第1項所述之字串鏈結方法，其中運算兩兩字串間之相似度值，更包括：使用動態規劃演算法，來找出兩字串間之最長共通字串；以及將該共通字串長度進行正規化以獲至該相似度值。
3. 如申請專利範圍第2項所述之字串鏈結方法，其中將此共通字串長度進行正規化更包括：計算兩字串之平均字串長度；以及將此共通字串長度除以該平均字串長度。

4. 如申請專利範圍第1項所述之字串鏈結方法，更包括計算各判別字串出現次數。
5. 如申請專利範圍第1項所述之字串鏈結方法，更包括定義該簡稱對照表。
6. 一種字串鏈結方法，係用於分類複數篇技術文件，其中每一篇技術文件具有一判別字串，包括：輸入欲進行分類之判別字串；根據所輸入之判別字串，計算各判別字串所對應之技術文件數目；使用一簡稱對照表，將該些判別字串中特定之簡寫還原；運算兩兩判別字串間之相似度值；將兩判別字串間相似度值高於一臨界值之兩判別字串判定為同一字串，並合併該兩判別字串對應之技術文件；以及持續進行相似度比對直至兩兩判別字串間之相似度均低於臨界值。
7. 如申請專利範圍第6項所述之字串鏈結方法，其中運算兩兩判別字串間之相似度值，更包括：使用動態規劃演算法，來找出兩判別字串間之最長共通字串；以及將該共通字串長度進行正規化以獲至該相似度值。
8. 如申請專利範圍第7項所述之字串鏈結方法，其中將此共通字串長度進行正規化更包括：計算兩判別字串之平均字串長度；以及將該共通字串長度除以該平均字串長度。
9. 如申請專利範圍第6項所述之字串鏈結方法，其中合併該兩判別字串對應之技術文件，更包括具有較少技術文件數目之對應判別字串會被消滅，而存留具有較多技術文件數目之對應判別字串。
10. 如申請專利範圍第6項所述之字串鏈結方法，更包括定義該簡稱對照表。

八、圖式：



第 1 圖

第1圖