



(21)申請案號：106131720

(22)申請日：中華民國 106 (2017) 年 09 月 15 日

(51)Int. Cl.：G06F16/00 (2019.01)G06Q40/08 (2012.01)

(30)優先權：2016/11/25 中國大陸201611055619.8

(71)申請人：開曼群島商創新先進技術有限公司 (開曼群島) ADVANCED NEW TECHNOLOGIES CO., LTD. (KY)
開曼群島

(72)發明人：孫清清 (CN)

(74)代理人：林志剛

(56)參考文獻：
CN 101727464B
US 2004/0024760A1
US 2013/0282645A1

審查人員：林育弘

申請專利範圍項數：22 項 圖式數：5 共 44 頁

(54)名稱
名稱匹配方法及裝置

(57)摘要

本發明實施例公開了一種名稱匹配方法及裝置。所述方法包括：獲得待匹配名稱；確定用於與所述待匹配名稱匹配的標準名稱集合；對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同；根據檢測結果，確定所述待匹配名稱的匹配結果。利用本發明實施例，可以提高匹配精確度，降低風險控制系統的誤報機率。

指定代表圖：

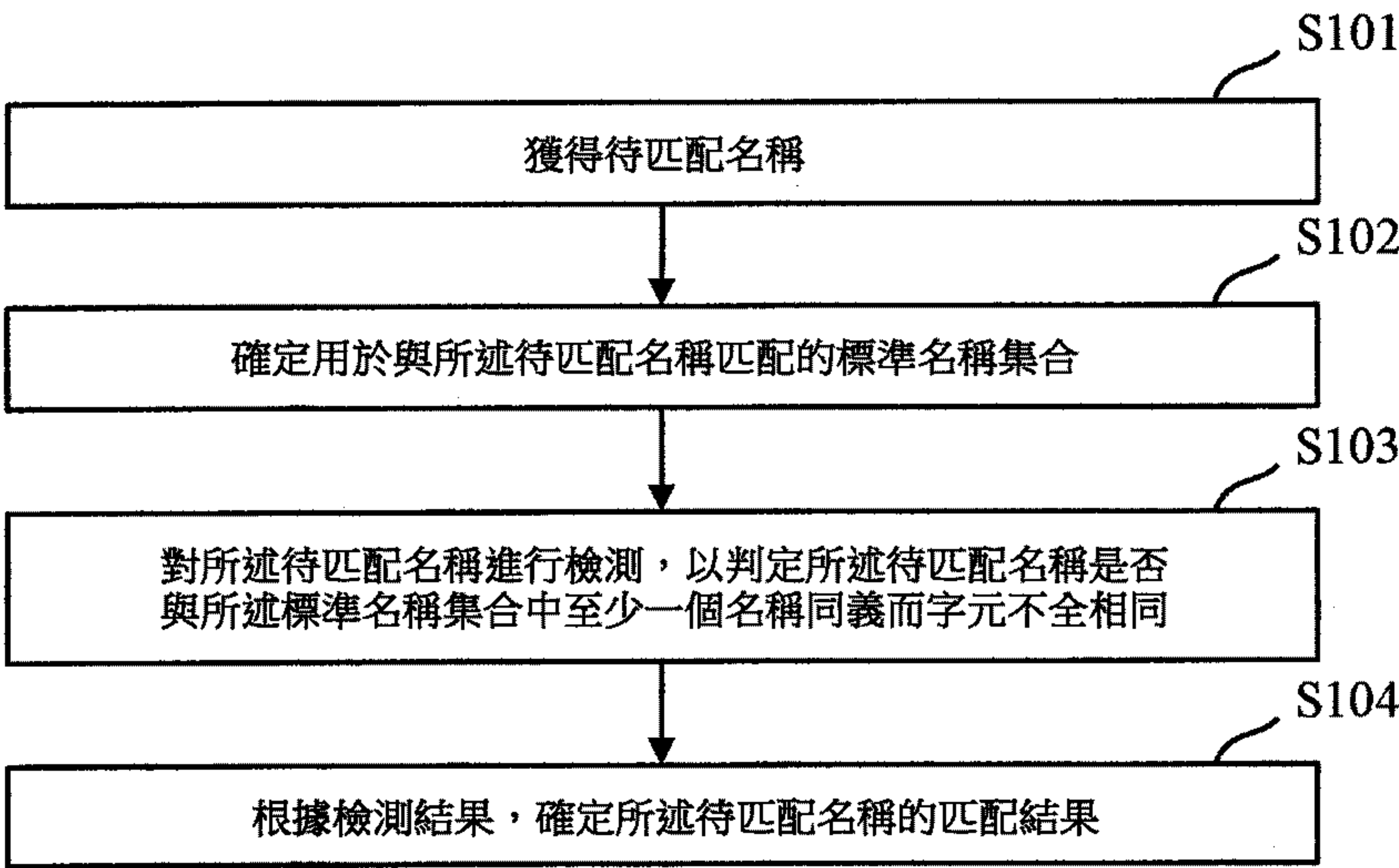


圖 1



I724237

【發明摘要】

【中文發明名稱】

名稱匹配方法及裝置

【中文】

本發明實施例公開了一種名稱匹配方法及裝置。所述方法包括：獲得待匹配名稱；確定用於與所述待匹配名稱匹配的標準名稱集合；對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同；根據檢測結果，確定所述待匹配名稱的匹配結果。利用本發明實施例，可以提高匹配精確度，降低風險控制系統的誤報機率。

【指定代表圖】第(1)圖。
【代表圖之符號簡單說明】無

【特徵化學式】無

【發明說明書】

【中文發明名稱】

名稱匹配方法及裝置

【技術領域】

本發明涉及電腦軟體技術領域，尤其涉及一種名稱匹配方法及裝置。

【先前技術】

人名匹配是風險控制領域非常重要的技術。例如，風險控制系統在黑名單中記錄已確定的各非法用戶的人名，之後在進行風險控制時，針對當前進行業務的每個使用者，通過掃描將每個使用者的人名與黑名單中的各人名進行匹配，若匹配成功，則可以認為該用戶是非法用戶並拒絕其業務，以防止風險。

人名匹配具體可以分為人名精確匹配和人名模糊匹配，相比而言，人名模糊匹配更具技術難點，因為難以把握合適的模糊程度。

在現有技術中，通常採用字串匹配演算法進行人名模糊匹配，則字串匹配度閾值決定了模糊程度，而字串匹配度閾值全憑經驗設定，為了減少遺漏，字串匹配度閾值往往設定得較低，從而容易導致匹配精確度較低，風險控制系統的誤報機率偏高。

【發明內容】

本發明實施例提供一種名稱匹配方法及裝置，用以解決如下技術問題：現有技術中採用字串匹配演算法進行人名模糊匹配的匹配精確度較低，系統誤報機率偏高。

為解決上述技術問題，本發明實施例是這樣實現的：

本發明實施例提供的一種名稱匹配方法，包括：

獲得待匹配名稱；

確定用於與所述待匹配名稱匹配的標準名稱集合；

對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同；

根據檢測結果，確定所述待匹配名稱的匹配結果。

本發明實施例提供的一種名稱匹配裝置，包括：

獲得模組，獲得待匹配名稱；

確定模組，確定用於與所述待匹配名稱匹配的標準名稱集合；

檢測模組，對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同；

匹配模組，根據檢測結果，確定所述待匹配名稱的匹配結果。

本發明實施例採用的上述至少一個技術方案能夠達到以下有益效果：所述名稱可以包括人名，實際應用中的待匹配人名可能由於資料錄入時的時效性、不確定性以及人

名的多變性，導致待匹配人名與實際人名有出入，這也是進行模糊匹配的原因，本發明的方案針對該原因，對待匹配人名進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同，並根據檢測結果確定人名匹配結果，相比於現有技術中僅局限於通過憑經驗設定的字串匹配度閾值控制模糊程度，更有利於提高對模糊程度的控制準確程度，可以提高匹配精確度，降低風險控制系統的誤報機率，因此，可以部分或全部地減少現有技術中的問題。

【圖式簡單說明】

為了更清楚地說明本發明實施例或現有技術中的技術方案，下面將對實施例或現有技術描述中所需要使用的圖式作簡單地介紹，顯而易見地，下面描述中的圖式僅僅是本發明中記載的一些實施例，對於本領域普通技術人員來講，在不付出創造性勞動性的前提下，還可以根據這些圖式獲得其他的圖式。

圖 1 為本發明實施例提供的一種名稱匹配方法的流程示意圖；

圖 2 為本發明實施例提供的一種實際應用場景下，名稱匹配方法中初篩的一種具體實施方案的流程簡圖；

圖 3 為本發明實施例提供的一種實際應用場景下，名稱匹配方法中集成演算法模糊匹配的一種具體實施方案的流程簡圖；

圖 4 為本發明實施例提供的一種實際應用場景下，名稱匹配方法的一種具體實施方案的流程簡圖；

圖 5 為本發明實施例提供的對應於圖 1 的一種名稱匹配裝置的結構示意圖。

【實施方式】

本發明實施例提供一種名稱匹配方法及裝置。

為了使本技術領域的人員更好地理解本發明中的技術方案，下面將結合本發明實施例中的圖式，對本發明實施例中的技術方案進行清楚、完整地描述，顯然，所描述的實施例僅僅是本發明一部分實施例，而不是全部的實施例。基於本發明中的實施例，本領域普通技術人員在沒有作出創造性勞動前提下所獲得的所有其他實施例，都應當屬於本發明保護的範圍。

如前所述，實際應用中的待匹配人名可能由於資料錄入時的時效性、不確定性以及人名的多變性，導致待匹配人名與實際人名有出入（主要指人名的“形（也即，字元）”發生變化）。為了便於理解，以英文人名為例，英文人名常見變化類型及實例，如下表 1 所示。

表 1

英文人名常見變化類型	人名變化類型的實例
錯拼	Jaxson、Jakson、Jaxon 本來是想表示同一個人名；SMYTH、SMITH 本來是想表示同一個人名；拼寫錯誤
語音拼寫差異	Michel、Michal、Miguel 本來是想表示同一個人名；發音類似導致的錯拼
暱稱	Mike、Mick、Mikey 表示同一人名的暱稱
稱謂詞	Dr.、Mr.等是對人的稱謂詞，和人名一起出現時不影響該人名的義
同義	Mohamed、Mohammed 和 Mohammad 出現在人名中時同義
縮寫	人名 ROB 是 ROBBIN 縮寫詞，二者同義；人名 BOB 是 BOBBY 的縮寫詞，二者同義；
多國語言差異	俄語 BORISOV 與 BORISOVNA 不同義，西班牙語 JUANLTA 和 JUAN 同義

其中，若只採用字串匹配的方法匹配人名，可能對於個別字母“錯拼”的變化類型尚能夠嘗試匹配（實際上也是靠誤打誤撞，沒有完美的根據），但是對於其他變化類型，匹配精確度很低。

而本發明的方案針對上述其他變化類型，可以進行諸如縮寫檢測、稱謂詞檢測、多語言檢測、別稱檢測等特定檢測，從而可以比較全面地考慮人名匹配時同義而字元不全相同（也即，同義而不同形）情況，進而可以提高匹配精確度。需要說明的是，“不同形”可以指由於錯拼而導致“有誤”的不同形，但在下面的實施例中更主要是指由於上述其他變化類型所導致的“合理無誤”的不同形。

本發明的方案不僅適用於人名匹配，也適用於人名以外的其他名稱的匹配，比如，地名、物名等。

下面對本發明的方案進行詳細說明。

圖 1 為本發明實施例提供的一種名稱匹配方法的流程示意圖。執行該流程的程式可搭載的設備包括但不限於：個人電腦、大中型電腦、電腦叢集、手機、平板電腦、智慧型可穿戴設備、車機，等等。該流程通常可以在風險控制領域使用，具體由風險控制系統或相關系統執行。

圖 1 中的流程可以包括以下步驟：

S101：獲得待匹配名稱。

在本發明實施例中，對待匹配名稱所屬的具體語言並不作限定，可以是英文、俄文、西班牙文等，也可以是中文等。為了便於描述，下面的實施例主要以待匹配名稱所屬的語言是英文為例進行說明。

S102：確定用於與所述待匹配名稱匹配的標準名稱集合。

在本發明實施例中，標準名稱集合可以是從一個更大的名稱集合中篩選出來的子集，也可以直接是所述更大的名稱集合本身，為了便於描述，可以將此處的篩選稱為“初篩”。比如，在背景技術的場景下，所述更大的名稱集合可以是風險控制系統所掌握的黑名單。

在前一種情況下，子集可以是只包含與待匹配名稱相似的名稱的集合，此處的“相似”可以不是那麼嚴格，因為後續還有一系列進一步地確定相似性的操作，相當於後續

還要進行“細節”。通過初篩可以快速縮小匹配範圍，進而可以減少後續細節的工作量，提高細節的針對性，有利於提高本發明的方案的效率。

S103：對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

在本發明實施例中，所要檢測判定的“同義而字元不全相同”主要是由表 1 中的一種或多種變化類型導致的，所述檢測具體可以包括：縮寫檢測、稱謂詞檢測、多語言檢測、別稱檢測等檢測中的至少一種，後面會一一詳細說明。

在本發明實施例中，當所述檢測包括多種時，可以按照一定的順序依次執行多種檢測，若在檢測過程中已經可以確定待匹配名稱的匹配結果，也可以不執行剩餘的檢測。當然，為了提高執行效率，也可以並行地執行多種檢測，再匯總檢測結果。

S104：根據檢測結果，確定所述待匹配名稱的匹配結果。

在本發明實施例中，通過執行步驟 **S103**，若判定待匹配名稱與標準名稱集合中至少一個名稱同義而字元不全相同，則根據所述情況可以直接確定待匹配名稱的匹配結果，在這種情況下，步驟 **S103** 中的檢測過程實際上也就是對待匹配名稱的整個匹配過程。

而若判定待匹配名稱不滿足：與所述標準名稱集合中

至少一個名稱同義而字元不全相同；則可以進一步地採用其他匹配方式對待匹配名稱進行匹配，以確定待匹配名稱的匹配結果。

通過圖 1 的方法，所述名稱可以包括人名，實際應用中的待匹配人名可能由於資料錄入時的時效性、不確定性以及人名的多變性，導致待匹配人名與實際人名有出入，這也是進行模糊匹配的原因，本發明的方案針對該原因，對待匹配人名進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同，並根據檢測結果確定人名匹配結果，相比於現有技術中僅局限於通過憑經驗設定的字串匹配度閾值控制模糊程度，更有利於提高對模糊程度的控制準確程度，可以提高匹配精確度，降低風險控制系統的誤報機率，因此，可以部分或全部地減少現有技術中的問題。

基於圖 1 的方法，本發明實施例還提供了該方法的一些具體實施方案，以及擴展方案，下面進行說明。

在本發明實施例中，不同的待匹配名稱的複雜程度可能不同，包含的資訊也可能不同，對於某些信息量過於小或者資訊特徵過於簡單的待匹配名稱，即使對其進行匹配，所獲得的匹配結果的價值也難以達到預期。比如，過於簡單通用的名稱如英文人名“Jim”、“Jimmy”、“David”、“John”、“Mike”等，即使匹配成功也難以特定到一個人。

為了防止這種情況導致用於名稱匹配的處理資源白白浪費，可以在獲得待匹配名稱後先對其進行一次過濾，以

判決是否繼續匹配。另外，若待匹配名稱包含在白名單上，則也存在類似問題，也可以採用這種方式進行處理。

具體地，對於步驟 S101 和 S102，所述獲得待匹配名稱後，確定用於與所述待匹配名稱匹配的標準名稱集合前，還可以執行：獲得預定的無需匹配名稱集合；確定所述待匹配名稱是否包含在所述無需匹配名單集合中；若是，則繼續執行後續步驟；否則，可以不對所述待匹配名稱進行匹配。

在本發明實施例中，以步驟 S102 的一種具體實施方案（對應於上述“初篩”的情況）為例展開說明。對於步驟 S102，所述確定用於與所述待匹配名稱匹配的標準名稱集合，具體可以包括：確定可用於與所述待匹配名稱匹配的第一名稱集合；通過將所述待匹配名稱包含的各詞與所述第一名稱集合中名稱包含的各詞進行相似性匹配，確定用於與所述待匹配名稱匹配的標準名稱集合。

如何進行相似性匹配也有多種具體實施方案，可以進行對待匹配名稱進行分詞匹配，也可以對待匹配進行全文本匹配，等等。

以分詞匹配為例，所述通過將所述待匹配名稱包含的各詞與所述第一名稱集合中名稱包含的各詞進行相似性匹配，確定用於與所述待匹配名稱匹配的標準名稱集合，具體可以包括：獲得所述第一名稱集合包含的各名稱的索引，所述名稱的索引為該名稱包含的任一詞；以及對所述待匹配名稱進行分詞，獲得所述待匹配名稱包含的各詞；

分別將對所述待匹配名稱包含的各詞，與各所述索引進行相似性匹配，獲得由匹配成功的各索引所索引的名稱構成的所述第一名稱集合的子集；根據各所述子集，確定用於與所述待匹配名稱匹配的標準名稱集合。

上例中的索引是預先建立的，基於索引進行分詞匹配的優點是：可以有效加快在匹配過程中對集合中所需的名稱的獲取速度。若不基於索引，仍可以實現分詞匹配（比如，在儲存集合的資料表中直接用 **Select** 語句查詢所需的名稱以用於分詞匹配等），只是效率上可能受到影響。

進一步地，所述分別將對所述待匹配名稱包含的各詞，與各所述索引進行相似性匹配，具體可以包括：採用字串匹配演算法，分別將對所述待匹配名稱包含的各詞，與各所述索引進行相似性匹配；所述字串匹配演算法可以包括一種或多種，比如，首碼樹匹配演算法、字典樹匹配演算法、字串相似度匹配演算法、發音相似匹配演算法等。這裡，採用字串匹配演算法只是一種較優的方式，也可以採用文本匹配演算法等其他可以用於實現相似性匹配的演算法。

所述根據各所述子集確定用於與所述待匹配名稱匹配的標準名稱集合，也可以有多種具體實施方案。比如，若採用 N 種字串匹配演算法，對於分詞得到的 M 個詞中的每個詞，分別採用每種字串匹配演算法將該詞與各索引進行匹配，對應得到 N 個子集，然後取 N 個子集的並集；共得到 M 個並集，取在 M 個並集中出現總次數超過設定

閾值的各名稱，確定為標準名稱集合。又比如，得到 N 個子集，然後取 N 個子集的交集確定為標準名稱集合。

上面對“初篩”過程進行了說明，為了便於理解，本發明實施例還提供了一種實際應用場景下，名稱匹配方法中初篩的一種具體實施方案的流程簡圖，如圖 2 所示。

在該實際應用場景下，假定第一名稱集合為某英文人名名單，預先設定了該名單中各名稱的索引（通過索引還可以索引得到對應名稱以外的相關資訊），索引具體為其對應的名稱中的詞，各索引包含在以索引對應的詞為主鍵建立的索引表中，待匹配名稱為英文人名“Kit Wai Jackson Wong”，以空格作為分隔符號對該人名進行分詞，分詞結果如下表 2 所示。

表 2

待匹配人名	Kit Wai Jackson Wong			
分詞後	詞 1: Kit	詞 2: Wai	詞 3: Jackson	詞 4: Wong

在圖 2 中，M=4，分詞結果具體為 {Kit，Wai，Jackson，Wong}，採用首碼樹匹配演算法、字典樹匹配演算法、字串相似度匹配演算法（比如 Simstring 演算法等）、發音相似匹配演算法（比如 metaphone 演算法等）分別對分詞得到的每個詞與各索引分別進行匹配，輸出單個詞通過這四種匹配演算法得到的索引對應的名稱集合，匹配結果 1、2、3、4；

將每個詞的匹配結果取並集，得到每個詞的綜合匹配結果；

取至少包含在兩個綜合匹配結果中的名稱構成集合，作為產生的初篩結果。

在本發明實施例中，確定標準名稱集合後，可以對待匹配名稱進行檢測。不過，在檢測之前，還可以進行一項或多項預處理，預處理有利於提高後續檢測結果的可靠性。所述預處理可以包括對齊處理、大小寫統一處理、簡繁體處理等。

以對齊處理為例。對於步驟 S103，所述對所述待匹配名稱進行檢測前，還可以執行：根據所述標準名稱集合中名稱包含的各詞與所述待匹配名稱包含的各詞的相似度，對所述標準名稱集合中名稱包含的各詞與所述待匹配名稱包含的各詞進行對齊處理。

在實際應用中，基於相似度進行對齊處理的具體原理可以是：基於相似度最大化原則，即將相似度最大的詞所在的名稱位置對齊。

例如，假定待匹配名稱為“Kate Lee Smith”，標準名稱集合中的某名稱為“Smith Catherine Lee”對齊結果如下表 3 所示。

表 3

待匹配名稱	Kate	Lee	Smith
標準名稱集合中對齊處理後的某名稱	Catherine	Lee	Smith

進一步地，在進行了對齊處理的情況下，對於步驟 S103，所述對所述待匹配名稱進行檢測，具體可以包括：根據對齊後的所述標準名稱集合，對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

在本發明實施例中，步驟 S103 中的檢測是提高人名模糊匹配的匹配準確度的關鍵部分。具體地，對於步驟 S103，所述對所述待匹配名稱進行檢測，具體包括：

對所述待匹配名稱進行縮寫檢測、稱謂詞檢測、多語言檢測、別稱檢測中的至少一種，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。下面分別對這幾種檢測進行說明。

在本發明實施例中，縮寫檢測中最常見的一種是首字母縮寫檢測，除此之外還有部分詞省略縮寫檢測等。在一種具體實施方式下，所述對所述待匹配名稱進行縮寫檢測，具體可以包括：獲得預定的縮寫對照組合資料，每個所述縮寫對照組合反映至少一個詞與其縮寫詞的縮寫對應關係；根據所述縮寫對照組合資料，檢測所述待匹配名稱包含的詞與所述標準名稱集合中的名稱包含的詞是否具有所述縮寫對應關係；根據檢測結果，判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

例如，若預先確定了英文人名“Ben Williams”可縮寫為“B. Williams”，則可以將“Ben Williams”與“B. Williams”

作為一個縮寫對照組合，若檢測出待匹配名稱與標準名稱集合中的某個名稱即為該縮寫對照組合，則可以確定存在同義而字元不全相同的情況。

在本發明實施例中，所述對所述待匹配名稱進行稱謂詞檢測，具體可以包括：獲得預定的稱謂詞資料；根據所述稱謂詞資料，檢測所述待匹配名稱是否包含所述稱謂詞，所述稱謂詞被認為不影響包含該稱謂詞的名稱之義；根據檢測結果，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

就人名而言，稱謂詞一般是通用尊稱（比如，**Mr.**或**Miss.**等）、頭銜（**Dr.**、**Prof.**等）等附加於至少部分原始人名的詞。在人名匹配的環境下，稱謂詞並不會影響其所對應的所述至少部分原始人名之義。因此，若確定待匹配名稱中包含有稱謂詞，且除了稱謂詞的其他部分與標準名稱集合中的某個名稱可以匹配，則可以判定待匹配名稱與標準名稱集合中至少一個名稱同義而字元不全相同。

在本發明實施例中，所述對所述待匹配名稱進行別稱檢測，具體可以包括：獲得預定的別稱對照組合資料，每個所述縮寫對照組合反映至少一個詞與其別稱的別稱對應關係；根據所述別稱對照組合資料，檢測所述待匹配名稱包含的詞與所述標準名稱集合中的名稱包含的詞是否具有所述別稱對應關係；根據檢測結果，判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

在實際應用中，別稱可以包括其對應的名稱的暱稱（比如，表 1 中的 Mick 是 Mikey 的暱稱等）、或其對應的名稱在不同領域下的同義名稱。對於後者，不同領域具體可以是不同地域（比如，不同國家、不同省份等）、不同語言（比如，不同國家的語言、不同民族的語言等）、不同的行業等。

相應地，對所述待匹配名稱進行暱稱檢測，和/或，對所述待匹配名稱進行不同領域下的同義名稱檢測。

別稱也不會影響其對應的名稱之義，因此，若確定待匹配名稱為標準名稱集合中的某個名稱的別稱，則可以判定待匹配名稱與標準名稱集合中至少一個名稱同義而字元不全相同。

需要說明的是，本發明對上述的對照組合資料、稱謂詞資料的具體儲存形式並不做限定，常用的一種方式是儲存在對應的資料表中，需要使用時從資料庫中讀取。

在本發明實施例中，所述對所述待匹配名稱進行多語言檢測，具體可以包括：確定所述待匹配名稱對應的語言；獲得所述語言和/或其他語言的拼寫變形同義規則和/或拼寫變形異義規則；根據所述拼寫變形同義規則和/或拼寫變形異義規則，對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

多語言檢測主要是針對諸如以下這種情況：比如，英文“Pooh”在德文中拼寫為“puh”，兩者在人名中出現是屬

於同義。

需要說明的是，在前面進行的各項檢測中，可能會使用到用於計算字串匹配度的演算法。

在本發明實施例中，所述根據檢測結果，確定所述待匹配名稱的匹配結果，具體可以包括：若判定所述待匹配名稱與所述標準名稱集合中至少一個名稱同義而字元不全相同，將所述至少一個名稱確定為所述待匹配名稱的匹配結果；否則（也即，通過上述檢測，未能匹配成功時），通過採用一種或多種相似度演算法對所述待匹配名稱與所述標準名稱集合中的名稱進行匹配，確定所述待匹配名稱的匹配結果。

在本發明實施例中，所述多種相似度演算法可以是基於不同維度的，從而有利於提高匹配結果的可靠性。根據這樣的思路，相似度演算法具體可以用於計算文本匹配度的演算法（比如，**n-gram** 演算法等）、用於計算語音匹配度的演算法（比如，**Phonex** 演算法等）、用於計算字串匹配度的演算法（比如，**Jaro-Winkler** 演算法等）等。

在採用了多種相似度演算法的情況下，可以通過對各相似度演算法對應的匹配結果的綜合衡量，得到綜合匹配結果，本發明對具體衡量方式並不做限定，常用的一種方式，是加權相加。

例如，採用 **n-gram** 演算法，演算法輸入為待匹配名稱包含的各詞與其對齊位置的各詞，演算法輸出為對齊的每一對詞的文本匹配度，記作 **F1**；

採用 Jaro-Winkler 演算法，演算法輸入為待匹配名稱包含的各詞與其對齊位置的各詞，演算法輸出為對齊的每一對詞的字串匹配度，記作 F2；

採用 Phonex 演算法，演算法輸入為待匹配名稱包含的各詞與其對齊位置的各詞，演算法輸出為對齊的每一對詞的語音匹配度，記作 F3；

通過對文本匹配度、字串匹配度和語音匹配度加權相加得到對齊的每一對詞的綜合匹配度 F，如下公式所示：

$$F=w1*F1+ w2*F2+ w3*F3$$
，其中， $w1+w2+w3=1$ ；

根據各對詞的綜合匹配度 F，通過取平均值得到待匹配名稱與標準名稱集合中的名稱之間的匹配結果，比如，對於表 3 中的名稱對，得到的匹配結果如下表 4 所示。

表 4

入檢名字	Kate	Lee	Smith
名單名字	Catherine	Lee	Smith
綜合匹配度	0.872	1	1
匹配結果	$(0.872+1+1)/3=0.957$		

上面對初篩後的檢測以及匹配過程進行了說明，該過程中可能使用到多種匹配相關的演算法，本發明的方案在具體實施時，可以對這些可能使用的演算法進行集成，則該過程即是一種集成演算法模糊匹配的過程。

在本發明實施例中，在進行集成演算法模糊匹配後，

還可以進一步地進行一些後置規則過濾，比如，將匹配結果中的匹配度映射為文字描述資訊，或者根據具體場景對匹配度進行適當補償或者消減等。

根據上面的說明，更直觀地，本發明實施例還提供了一種實際應用場景下，名稱匹配方法中集成演算法模糊匹配的一種具體實施方案的流程簡圖，如圖 3 所示。在圖 3 中，各項檢測和各種方式的匹配度計算的執行順序只是示例，並非對本發明的限定。

在圖 3 中，若通過上述任一項檢測對待匹配名稱匹配成功，則可以直接輸出匹配結果，反之，則可以通過一種或多種方式的匹配度計算（比如，文本匹配度計算、語音匹配度計算、字串匹配度計算等），確定並輸出待匹配名稱的匹配結果。

當然，也可以完整地執行各項檢測和各種方式的匹配度計算，再綜合考慮各項檢測結果和各種方式的匹配度計算結果，確定待匹配字串的匹配結果。

進一步地，本發明實施例還提供了一種實際應用場景下，名稱匹配方法的一種具體實施方案的流程簡圖，如圖 4 所示。在圖 4 中，前置規則判決可以包括：判決待匹配名稱是否未包含在無需匹配名單集合中等，掃描名單索引即為上述的標準名稱集合中名稱的索引。

圖 3、圖 4 中的步驟在上面都進行了詳細說明，在此不贅述。

上面對本發明實施例提供的名稱匹配方法進行了說

明，基於同樣的發明思路，本發明實施例還提供了對應的裝置，如圖 5 所示。

圖 5 為本發明實施例提供的對應於圖 1 的一種名稱匹配裝置的結構示意圖，虛線表示可選的模組，所述裝置包括：

獲得模組 501，獲得待匹配名稱；

確定模組 502，確定用於與所述待匹配名稱匹配的標準名稱集合；

檢測模組 503，對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同；

匹配模組 504，根據檢測結果，確定所述待匹配名稱的匹配結果。

可選地，所述確定模組 502 確定用於與所述待匹配名稱匹配的標準名稱集合前，獲得預定的無需匹配名稱集合，確定所述待匹配名稱未包含在所述無需匹配名單集合中。

可選地，所述確定模組 502 確定用於與所述待匹配名稱匹配的標準名稱集合，具體包括：

所述確定模組 502 獲得可用於與所述待匹配名稱匹配的第一名稱集合，通過將所述待匹配名稱包含的各詞與所述第一名稱集合中名稱包含的各詞進行相似性匹配，確定用於與所述待匹配名稱匹配的標準名稱集合。

可選地，所述確定模組 502 通過將所述待匹配名稱包

含的各詞與所述第一名稱集合中名稱包含的各詞進行相似性匹配，確定用於與所述待匹配名稱匹配的標準名稱集合，具體包括：

所述確定模組 502 獲得所述第一名稱集合包含的各名稱的索引，所述名稱的索引為該名稱包含的任一詞，以及對所述待匹配名稱進行分詞，獲得所述待匹配名稱包含的各詞，分別將對所述待匹配名稱包含的各詞，與各所述索引進行相似性匹配，獲得由匹配成功的各索引所索引的名稱構成的所述第一名稱集合的子集，根據各所述子集，確定用於與所述待匹配名稱匹配的標準名稱集合。

可選地，所述確定模組 502 分別將對所述待匹配名稱包含的各詞，與各所述索引進行相似性匹配，具體包括：

所述確定模組 502 採用字串匹配演算法，分別將對所述待匹配名稱包含的各詞，與各所述索引進行相似性匹配，所述字串匹配演算法包括以下至少一種：首碼樹匹配演算法、字典樹匹配演算法、字串相似度匹配演算法、發音相似匹配演算法。

可選地，所述裝置還包括：

對齊模組 505，在所述檢測模組 503 對所述待匹配名稱進行檢測前，根據所述標準名稱集合中名稱包含的各詞與所述待匹配名稱包含的各詞的相似度，對所述標準名稱集合中名稱包含的各詞與所述待匹配名稱包含的各詞進行對齊處理；

所述檢測模組 503 對所述待匹配名稱進行檢測，具體

包括：

所述檢測模組 503 根據對齊後的所述標準名稱集合，對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

可選地，所述檢測模組 503 對所述待匹配名稱進行檢測，具體包括：

所述檢測模組 503 對所述待匹配名稱進行縮寫檢測、稱謂詞檢測、多語言檢測、別稱檢測中的至少一種，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

可選地，所述檢測模組 503 對所述待匹配名稱進行縮寫檢測，具體包括：

所述檢測模組獲得預定的縮寫對照組合資料，每個所述縮寫對照組合反映至少一個詞與其縮寫詞的縮寫對應關係，根據所述縮寫對照組合資料，檢測所述待匹配名稱包含的詞與所述標準名稱集合中的名稱包含的詞是否具有所述縮寫對應關係，根據檢測結果，判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

可選地，所述檢測模組 503 對所述待匹配名稱進行稱謂詞檢測，具體包括：

所述檢測模組 503 獲得預定的稱謂詞資料，根據所述稱謂詞資料，檢測所述待匹配名稱是否包含所述稱謂詞，

所述稱謂詞被認為不影響包含該稱謂詞的名稱之義，根據檢測結果，判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

可選地，所述別稱包括其對應的名稱的暱稱、或其對應的名稱在不同領域下的同義名稱；

所述檢測模組 503 對所述待匹配名稱進行別稱檢測，具體包括：

所述檢測模組 503 對所述待匹配名稱進行暱稱檢測，和/或，對所述待匹配名稱進行不同領域下的同義名稱檢測。

可選地，所述檢測模組 503 對所述待匹配名稱進行多語言檢測，具體包括：

所述檢測模組 503 確定所述待匹配名稱對應的語言，獲得所述語言的拼寫變形同義規則和/或拼寫變形異義規則，根據所述拼寫變形同義規則和/或拼寫變形異義規則，對所述待匹配名稱進行檢測，以判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

可選地，所述檢測模組 503 對所述待匹配名稱進行別稱檢測，具體包括：

所述檢測模組 503 獲得預定的別稱對照組合資料，每個所述縮寫對照組合反映至少一個詞與其別稱的別稱對應關係，根據所述別稱對照組合資料，檢測所述待匹配名稱包含的詞與所述標準名稱集合中的名稱包含的詞是否具有

所述別稱對應關係，根據檢測結果，判定所述待匹配名稱是否與所述標準名稱集合中至少一個名稱同義而字元不全相同。

可選地，所述匹配模組 504 根據檢測結果，確定所述待匹配名稱的匹配結果，具體包括：

所述匹配模組 504 若所述檢測模組 503 判定所述待匹配名稱與所述標準名稱集合中至少一個名稱同義而字元不全相同，將所述至少一個名稱確定為所述待匹配名稱的匹配結果，否則，通過採用一種或多種相似度演算法對所述待匹配名稱與所述標準名稱集合中的名稱進行匹配，確定所述待匹配名稱的匹配結果。

可選地，所述相似度演算法包括以下至少一種：用於計算文本匹配度的演算法、用於計算語音匹配度的演算法、用於計算字串匹配度的演算法。

可選地，所述名稱為人名。

本發明實施例提供的裝置與方法是一一對應的，因此，裝置也具有與其對應的方法類似的有益技術效果，由於上面已經對方法的有益技術效果進行了詳細說明，因此，這裡不再贅述對應裝置的有益技術效果。

另外，在本發明對上述裝置與方法的具體應用場景並不做限定，除了背景技術中提到的風險控制領域（比如，反洗錢、用戶認證等具體領域）以外，其他任何可能需要用到名稱匹配技術的領域，本發明的方案均可能適用。

在 20 世紀 90 年代，對於一個技術的改進可以很明顯

地區分是硬體上的改進（例如，對二極體、電晶體、開關等電路結構的改進）還是軟體上的改進（對於方法流程的改進）。然而，隨著技術的發展，當今的很多方法流程的改進已經可以視為硬體電路結構的直接改進。設計人員幾乎都通過將改進的方法流程程式設計到硬體電路中來得到相應的硬體電路結構。因此，不能說一個方法流程的改進就不能用硬體實體模組來實現。例如，可程式設計邏輯裝置（**Programmable Logic Device, PLD**）（例如場可程式化閘陣列（**Field Programmable Gate Array, FPGA**））就是這樣一種積體電路，其邏輯功能由使用者對裝置程式設計來確定。由設計人員自行程式設計來把一個數位系統“集成”在一片 **PLD** 上，而不需要請晶片製造廠商來設計和製作專用的積體電路晶片。而且，如今，取代手工地製作積體電路晶片，這種程式設計也多半改用“邏輯編譯器（**logic compiler**）”軟體來實現，它與程式開發撰寫時所用的軟體編譯器相類似，而要編譯之前的原始代碼也得用特定的程式設計語言來撰寫，此稱之為硬體描述語言（**Hardware Description Language, HDL**），而 **HDL** 也並非僅有一種，而是有許多種，如 **ABEL**（**Advanced Boolean Expression Language**）、**AHDL**（**Altera Hardware Description Language**）、**Confluence**、**CUPL**（**Cornell University Programming Language**）、**HDCal**、**JHDL**（**Java Hardware Description Language**）、**Lava**、**Lola**、**MyHDL**、**PALASM**、**RHDL**（**Ruby Hardware Description**

Language) 等，目前最普遍使用的是 VHDL (Very-High-Speed Integrated Circuit Hardware Description Language) 與 Verilog。本領域技術人員也應該清楚，只需要將方法流程用上述幾種硬體描述語言稍作邏輯程式設計並程式化到積體電路中，就可以很容易得到實現該邏輯方法流程的硬體電路。

控制器可以按任何適當的方式實現，例如，控制器可以採取例如微處理器或處理器以及儲存可由該(微)處理器執行的電腦可讀程式碼(例如軟體或韌體)的電腦可讀介質、邏輯閘、開關、專用積體電路(Application Specific Integrated Circuit, ASIC)、可程式設計邏輯控制器和嵌入微控制器的形式，控制器的例子包括但不限於以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及 Silicone Labs C8051F320，記憶體控制器還可以被實現為記憶體的控制邏輯的一部分。本領域技術人員也知道，除了以純電腦可讀程式碼方式實現控制器以外，完全可以通過將方法步驟進行邏輯程式設計來使得控制器以邏輯閘、開關、專用積體電路、可程式化邏輯控制器和嵌入微控制器等的形式來實現相同功能。因此這種控制器可以被認為是一種硬體元件，而對其內包括的用於實現各種功能的裝置也可以視為硬體元件內的結構。或者甚至，可以將用於實現各種功能的裝置視為既可以是實現方法的軟體模組又可以是硬體元件內的結構。

上述實施例闡明的系統、裝置、模組或單元，具體可

以由電腦晶片或實體實現，或者由具有某種功能的產品來實現。一種典型的實現設備為電腦。具體的，電腦例如可以為個人電腦、筆記型電腦、蜂巢式電話、相機電話、智慧型型電話、個人數位助理、媒體播放機、導航設備、電子郵件設備、遊戲控制台、平板電腦、可穿戴設備或者這些設備中的任何設備的組合。

為了描述的方便，描述以上裝置時以功能分為各種單元分別描述。當然，在實施本發明時可以把各單元的功能在同一個或多個軟體和/或硬體中實現。

本領域內的技術人員應明白，本發明的實施例可提供為方法、系統、或電腦程式產品。因此，本發明可採用完全硬體實施例、完全軟體實施例、或結合軟體和硬體方面的實施例的形式。而且，本發明可採用在一個或多個其中包含有電腦可用程式碼的電腦可用儲存介質（包括但不限於磁碟記憶體、**CD-ROM**、光學記憶體等）上實施的電腦程式產品的形式。

本發明是參照根據本發明實施例的方法、設備（系統）、和電腦程式產品的流程圖和／或方塊圖來描述的。應理解可由電腦程式指令實現流程圖和／或方塊圖中的每一流程和／或方塊、以及流程圖和／或方塊圖中的流程和／或方塊的結合。可提供這些電腦程式指令到通用電腦、專用電腦、嵌入式處理機或其他可程式設計資料處理設備的處理器以產生一個機器，使得通過電腦或其他可程式設計資料處理設備的處理器執行的指令產生用於實現在流程

圖一個流程或多個流程和／或方塊圖一個方塊或多個方塊中指定的功能的裝置。

這些電腦程式指令也可儲存在能引導電腦或其他可程式設計資料處理設備以特定方式工作的電腦可讀記憶體中，使得儲存在該電腦可讀記憶體中的指令產生包括指令裝置的製造品，該指令裝置實現在流程圖一個流程或多個流程和／或方塊圖一個方塊或多個方塊中指定的功能。

這些電腦程式指令也可裝載到電腦或其他可程式設計資料處理設備上，使得在電腦或其他可程式設計設備上執行一系列操作步驟以產生電腦實現的處理，從而在電腦或其他可程式設計設備上執行的指令提供用於實現在流程圖一個流程或多個流程和／或方塊圖一個方塊或多個方塊中指定的功能的步驟。

在一個典型的配置中，計算設備包括一個或多個處理器(CPU)、輸入/輸出介面、網路介面和記憶體。

記憶體可能包括電腦可讀介質中的非永久性記憶體，隨機存取記憶體(RAM)和/或非揮發性記憶體等形式，如唯讀記憶體(ROM)或快閃記憶體(flash RAM)。記憶體是電腦可讀介質的示例。

電腦可讀介質包括永久性和非永久性、可移動和非可移動媒體可以由任何方法或技術來實現資訊儲存。資訊可以是電腦可讀指令、資料結構、程式的模組或其他資料。電腦的儲存介質的例子包括，但不限於相變記憶體(PRAM)、靜態隨機存取記憶體(SRAM)、動態隨機存取記

憶體(DRAM)、其他類型的隨機存取記憶體(RAM)、唯讀記憶體(ROM)、電可擦除可程式設計唯讀記憶體(EEPROM)、快閃記憶體或其他記憶體技術、唯讀光碟唯讀記憶體(CD-ROM)、數位多功能光碟(DVD)或其他光學儲存、磁盒式磁帶，磁帶磁磁片儲存或其他磁性存放裝置或任何其他非傳輸介質，可用於儲存可以被計算設備訪問的資訊。按照本文中的界定，電腦可讀介質不包括暫存電腦可讀媒體(transitory media)，如調變的資料信號和載波。

還需要說明的是，術語“包括”、“包含”或者其任何其他變體意在涵蓋非排他性的包含，從而使得包括一系列要素的過程、方法、商品或者設備不僅包括那些要素，而且還包括沒有明確列出的其他要素，或者是還包括為這種過程、方法、商品或者設備所固有的要素。在沒有更多限制的情況下，由語句“包括一個……”限定的要素，並不排除在包括所述要素的過程、方法、商品或者設備中還存在另外的相同要素。

本領域技術人員應明白，本發明的實施例可提供為方法、系統或電腦程式產品。因此，本發明可採用完全硬體實施例、完全軟體實施例或結合軟體和硬體方面的實施例的形式。而且，本發明可採用在一個或多個其中包含有電腦可用程式碼的電腦可用儲存介質（包括但不限於磁碟記憶體、CD-ROM、光學記憶體等）上實施的電腦程式產品的形式。

本發明可以在由電腦執行的電腦可執行指令的一般上

下文中描述，例如程式模組。一般地，程式模組包括執行特定任務或實現特定抽象資料類型的常式、程式、物件、元件、資料結構等等。也可以在分散式運算環境中實踐本發明，在這些分散式運算環境中，由通過通信網路而被連接的遠端處理設備來執行任務。在分散式運算環境中，程式模組可以位於包括存放裝置在內的本地和遠端電腦儲存介質中。

本說明書中的各個實施例均採用遞進的方式描述，各個實施例之間相同相似的部分互相參見即可，每個實施例重點說明的都是與其他實施例的不同之處。尤其，對於系統實施例而言，由於其基本相似於方法實施例，所以描述的比較簡單，相關之處參見方法實施例的部分說明即可。

以上所述僅為本發明的實施例而已，並不用於限制本發明。對於本領域技術人員來說，本發明可以有各種更改和變化。凡在本發明的精神和原理之內所作的任何修改、等同替換、改進等，均應包含在本發明的申請專利範圍之內。

【符號說明】

501：獲得模組

502：確定模組

503：檢測模組

504：匹配模組

505：對齊模組

【發明申請專利範圍】

【第1項】

一種名稱匹配方法，用於風險控制領域，以解決黑名單匹配問題，並以電子裝置執行，其特徵在於，包括：

獲得待匹配名稱；

確定用於與該待匹配名稱匹配的標準名稱集合；

對該待匹配名稱進行檢測，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同；

根據檢測結果，確定該待匹配名稱的匹配結果，具體包括：

若判定該待匹配名稱與該標準名稱集合中至少一個名稱同義而字元不全相同，將該至少一個名稱確定為該待匹配名稱的匹配結果；

否則，通過採用一種或多種相似度演算法對該待匹配名稱與該標準名稱集合中的名稱進行匹配，確定該待匹配名稱的匹配結果，

其中，該對該待匹配名稱進行檢測，具體包括：

對該待匹配名稱進行縮寫檢測、稱謂詞檢測、多語言檢測中的至少一種，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同，

其中，該相似度演算法包括以下至少一種：用於計算文本匹配度的演算法、用於計算語音匹配度的演算法、用於計算字串匹配度的演算法，

其中，該確定用於與該待匹配名稱匹配的標準名稱集

第 106131720 號

民國 110 年 2 月 3 日修正

合，具體包括：

確定可用於與該待匹配名稱匹配的第一名稱集合；

通過將該待匹配名稱包含的各詞與該第一名稱集合中名稱包含的各詞進行相似性匹配，確定用於與該待匹配名稱匹配的標準名稱集合。

【第2項】

如請求項1所述的方法，其中，該確定用於與該待匹配名稱匹配的標準名稱集合前，該方法還包括：

獲得預定的無需匹配名稱集合；

確定該待匹配名稱未包含在該無需匹配名單集合中。

【第3項】

如請求項1所述的方法，其中，該通過將該待匹配名稱包含的各詞與該第一名稱集合中名稱包含的各詞進行相似性匹配，確定用於與該待匹配名稱匹配的標準名稱集合，具體包括：

獲得該第一名稱集合包含的各名稱的索引，該名稱的索引為該名稱包含的任一詞；以及

對該待匹配名稱進行分詞，獲得該待匹配名稱包含的各詞；

分別將對該待匹配名稱包含的各詞，與各該索引進行相似性匹配，獲得由匹配成功的各索引所索引的名稱構成的該第一名稱集合的子集；

根據各該子集，確定用於與該待匹配名稱匹配的標準

名稱集合。

【第4項】

如請求項3所述的方法，其中，該分別將對該待匹配名稱包含的各詞，與各該索引進行相似性匹配，具體包括：

採用字串匹配演算法，分別將對該待匹配名稱包含的各詞，與各該索引進行相似性匹配；

該字串匹配演算法包括以下至少一種：首碼樹匹配演算法、字典樹匹配演算法、字串相似度匹配演算法、發音相似匹配演算法。

【第5項】

如請求項1所述的方法，其中，該對該待匹配名稱進行檢測前，該方法還包括：

根據該標準名稱集合中名稱包含的各詞與該待匹配名稱包含的各詞的相似度，對該標準名稱集合中名稱包含的各詞與該待匹配名稱包含的各詞進行對齊處理；
該對該待匹配名稱進行檢測，具體包括：

根據對齊後的該標準名稱集合，對該待匹配名稱進行檢測，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第6項】

如請求項1所述的方法，其中，該對該待匹配名稱進行縮寫檢測，具體包括：

獲得預定的縮寫對照組合資料，每個該縮寫對照組合

反映至少一個詞與其縮寫詞的縮寫對應關係；

根據該縮寫對照組合資料，檢測該待匹配名稱包含的詞與該標準名稱集中的名稱包含的詞是否具有該縮寫對應關係；

根據檢測結果，判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第7項】

如請求項1所述的方法，其中，該對該待匹配名稱進行稱謂詞檢測，具體包括：

獲得預定的稱謂詞資料；

根據該稱謂詞資料，檢測該待匹配名稱是否包含該稱謂詞，該稱謂詞被認為不影響包含該稱謂詞的名稱之義；

根據檢測結果，判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第8項】

如請求項1所述的方法，其中，該別稱包括其對應的名稱的暱稱、或其對應的名稱在不同領域下的同義名稱；

該對該待匹配名稱進行別稱檢測，具體包括：

對該待匹配名稱進行暱稱檢測，和/或，對該待匹配名稱進行不同領域下的同義名稱檢測。

【第9項】

如請求項1所述的方法，其中，該對該待匹配名稱進行多語言檢測，具體包括：

確定該待匹配名稱對應的語言；

獲得該語言和/或其他語言的拼寫變形同義規則和/或拼寫變形異義規則；

根據該拼寫變形同義規則和/或拼寫變形異義規則，對該待匹配名稱進行檢測，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第10項】

如請求項1所述的方法，其中，該對該待匹配名稱進行別稱檢測，具體包括：

獲得預定的別稱對照組合資料，每個該縮寫對照組合反映至少一個詞與其別稱的別稱對應關係；

根據該別稱對照組合資料，檢測該待匹配名稱包含的詞與該標準名稱集合中的名稱包含的詞是否具有該別稱對應關係；

根據檢測結果，判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第11項】

如請求項1～10中任一項所述的方法，其中，該名稱為人名。

【第12項】

一種名稱匹配裝置，用於風險控制領域，以解決黑名單匹配問題，其特徵在於，包括：

獲得模組，獲得待匹配名稱；

確定模組，確定用於與該待匹配名稱匹配的標準名稱集合；

檢測模組，對該待匹配名稱進行檢測，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同；

匹配模組，根據檢測結果，確定該待匹配名稱的匹配結果，具體包括：

該匹配模組若該檢測模組判定該待匹配名稱與該標準名稱集合中至少一個名稱同義而字元不全相同，將該至少一個名稱確定為該待匹配名稱的匹配結果，否則，通過採用一種或多種相似度演算法對該待匹配名稱與該標準名稱集合中的名稱進行匹配，確定該待匹配名稱的匹配結果，

其中，該檢測模組對該待匹配名稱進行檢測，具體包括：

該檢測模組對該待匹配名稱進行縮寫檢測、稱謂詞檢測、多語言檢測中的至少一種，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同，

其中，該相似度演算法包括以下至少一種：用於計算文本匹配度的演算法、用於計算語音匹配度的演算法、用於計算字串匹配度的演算法，

其中，該確定模組確定用於與該待匹配名稱匹配的標準名稱集合，具體包括：

該確定模組獲得可用於與該待匹配名稱匹配的第一名稱集合，通過將該待匹配名稱包含的各詞與該第一名

稱集合中名稱包含的各詞進行相似性匹配，確定用於與該待匹配名稱匹配的標準名稱集合。

【第13項】

如請求項12所述的裝置，其中，該確定模組確定用於與該待匹配名稱匹配的標準名稱集合前，獲得預定的無需匹配名稱集合，確定該待匹配名稱未包含在該無需匹配名單集合中。

【第14項】

如請求項12所述的裝置，其中，該確定模組通過將該待匹配名稱包含的各詞與該第一名稱集合中名稱包含的各詞進行相似性匹配，確定用於與該待匹配名稱匹配的標準名稱集合，具體包括：

該確定模組獲得該第一名稱集合包含的各名稱的索引，該名稱的索引為該名稱包含的任一詞，以及對該待匹配名稱進行分詞，獲得該待匹配名稱包含的各詞，分別將對該待匹配名稱包含的各詞，與各該索引進行相似性匹配，獲得由匹配成功的各索引所索引的名稱構成的該第一名稱集合的子集，根據各該子集，確定用於與該待匹配名稱匹配的標準名稱集合。

【第15項】

如請求項14所述的裝置，其中，該確定模組分別將對該待匹配名稱包含的各詞，與各該索引進行相似性匹配，具體包括：

該確定模組採用字串匹配演算法，分別將對該待匹配

第 106131720 號

民國 110 年 2 月 3 日修正

名稱包含的各詞，與各該索引進行相似性匹配，該字串匹配演算法包括以下至少一種：首碼樹匹配演算法、字典樹匹配演算法、字串相似度匹配演算法、發音相似匹配演算法。

【第16項】

如請求項12所述的裝置，其中，該裝置還包括：

對齊模組，在該檢測模組對該待匹配名稱進行檢測前，根據該標準名稱集合中名稱包含的各詞與該待匹配名稱包含的各詞的相似度，對該標準名稱集合中名稱包含的各詞與該待匹配名稱包含的各詞進行對齊處理；

該檢測模組對該待匹配名稱進行檢測，具體包括：

該檢測模組根據對齊後的該標準名稱集合，對該待匹配名稱進行檢測，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第17項】

如請求項12所述的裝置，其中，該檢測模組對該待匹配名稱進行縮寫檢測，具體包括：

該檢測模組獲得預定的縮寫對照組合資料，每個該縮寫對照組合反映至少一個詞與其縮寫詞的縮寫對應關係，根據該縮寫對照組合資料，檢測該待匹配名稱包含的詞與該標準名稱集合中的名稱包含的詞是否具有該縮寫對應關係，根據檢測結果，判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第18項】

如請求項12所述的裝置，其中，該檢測模組對該待匹配名稱進行稱謂詞檢測，具體包括：

該檢測模組獲得預定的稱謂詞資料，根據該稱謂詞資料，檢測該待匹配名稱是否包含該稱謂詞，該稱謂詞被認為不影響包含該稱謂詞的名稱之義，根據檢測結果，判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第19項】

如請求項12所述的裝置，其中，該別稱包括其對應的名稱的暱稱、或其對應的名稱在不同領域下的同義名稱；

該檢測模組對該待匹配名稱進行別稱檢測，具體包括：

該檢測模組對該待匹配名稱進行暱稱檢測，和/或，對該待匹配名稱進行不同領域下的同義名稱檢測。

【第20項】

如請求項12所述的裝置，其中，該檢測模組對該待匹配名稱進行多語言檢測，具體包括：

該檢測模組確定該待匹配名稱對應的語言，獲得該語言和/或其他語言的拼寫變形同義規則和/或拼寫變形異義規則，根據該拼寫變形同義規則和/或拼寫變形異義規則，對該待匹配名稱進行檢測，以判定該待匹配名稱是否與該標準名稱集合中至少一個名稱同義而字元不全相同。

【第21項】

如請求項12所述的裝置，其中，該檢測模組對該待匹

配名稱進行別稱檢測，具體包括：

該檢測模組獲得預定的別稱對照組合資料，每個該縮寫對照組合反映至少一個詞與其別稱的別稱對應關係，根據該別稱對照組合資料，檢測該待匹配名稱包含的詞與該標準名稱集中的名稱包含的詞是否具有該別稱對應關係，根據檢測結果，判定該待匹配名稱是否與該標準名稱集中至少一個名稱同義而字元不全相同。

【第22項】

如請求項12～21中任一項所述的裝置，其中，該名稱為人名。

【發明圖式】

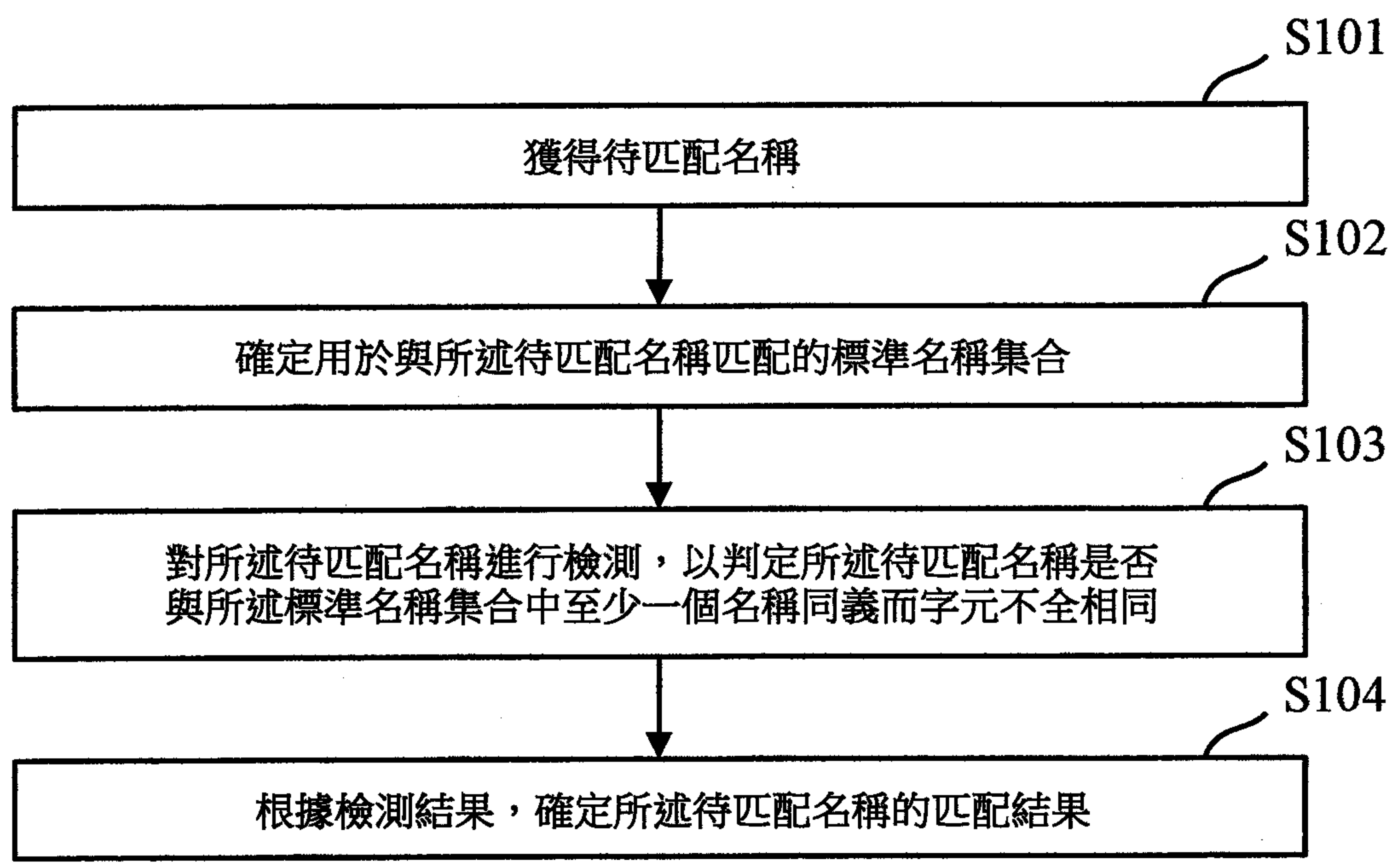


圖 1

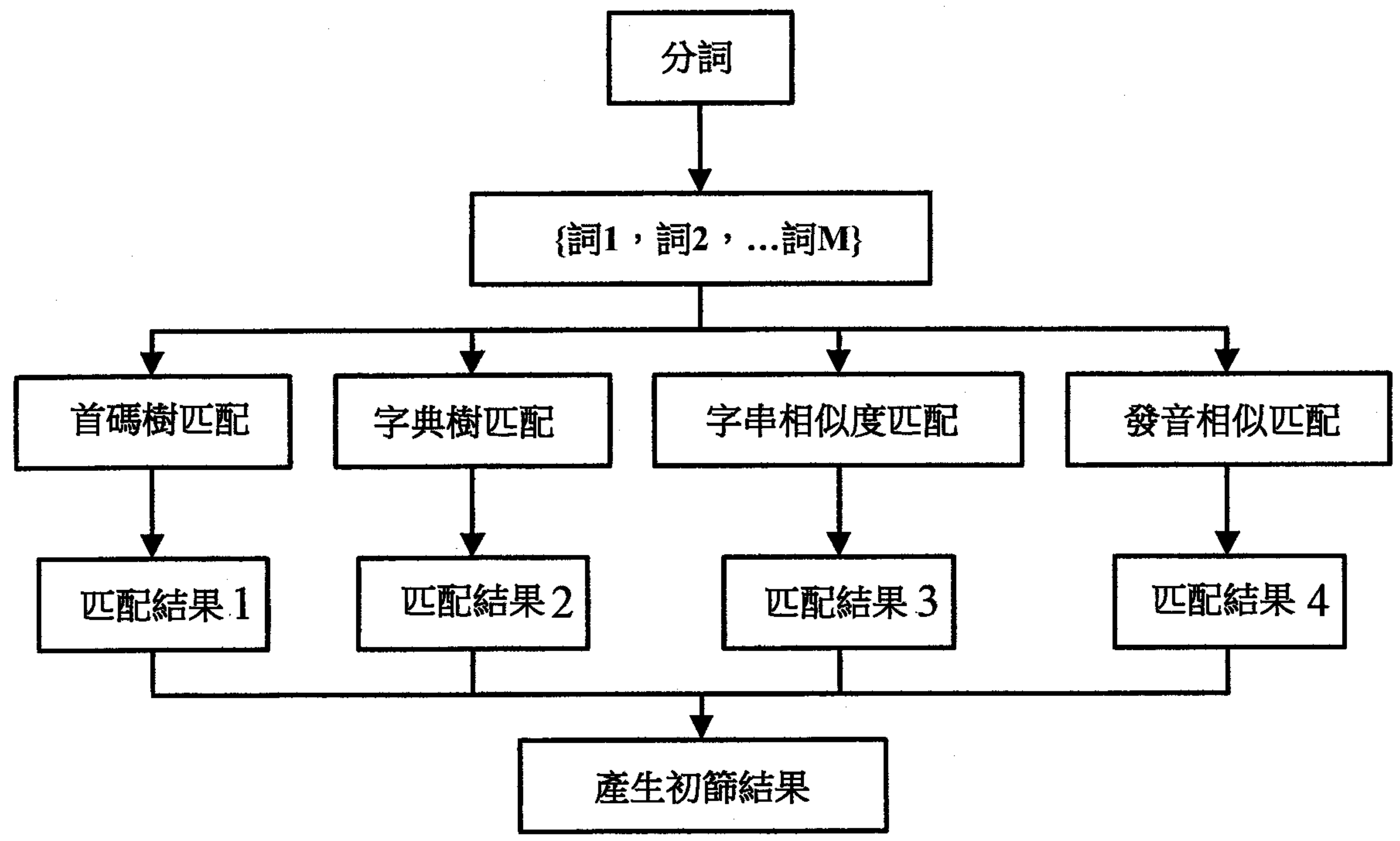


圖 2

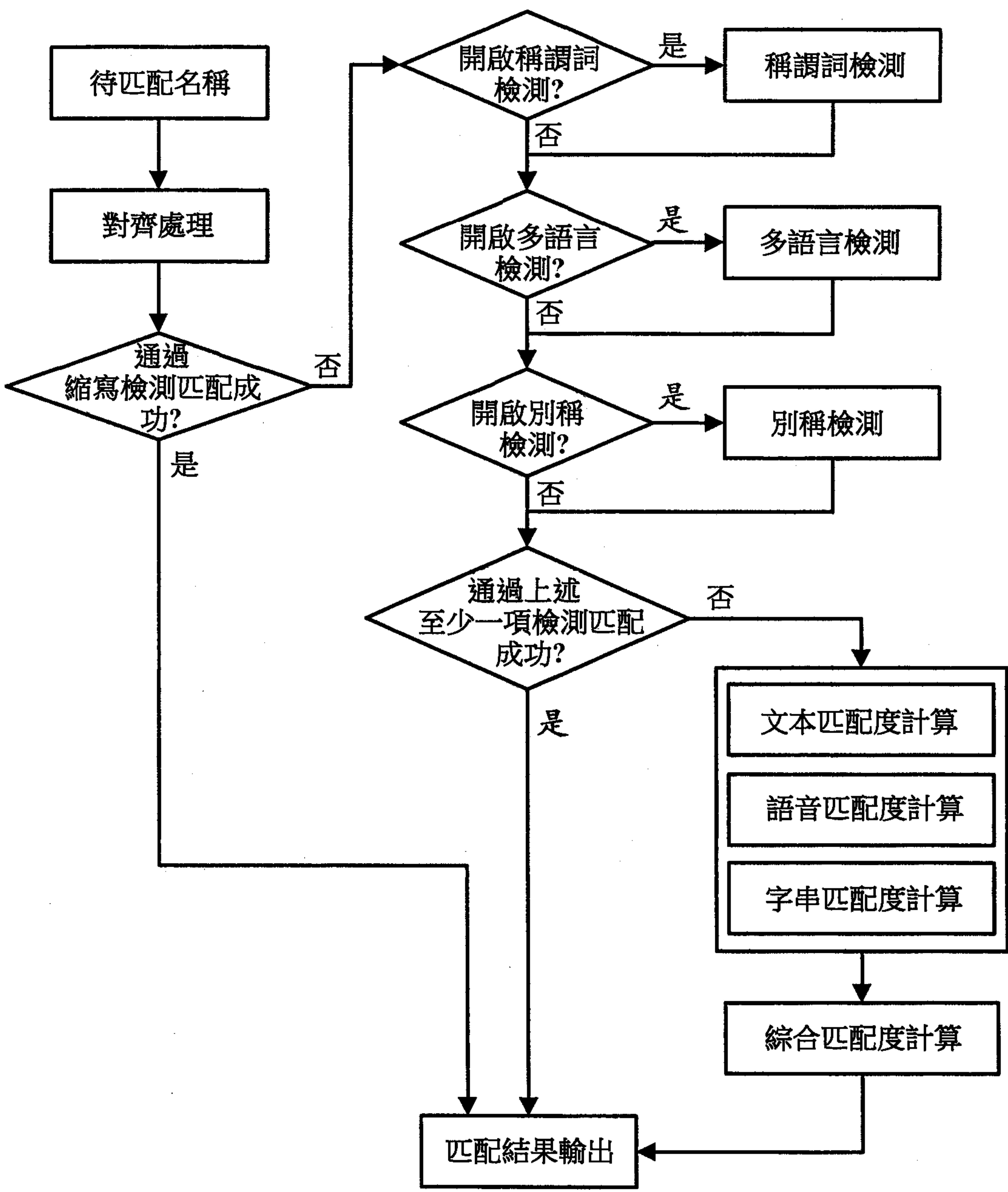


圖 3

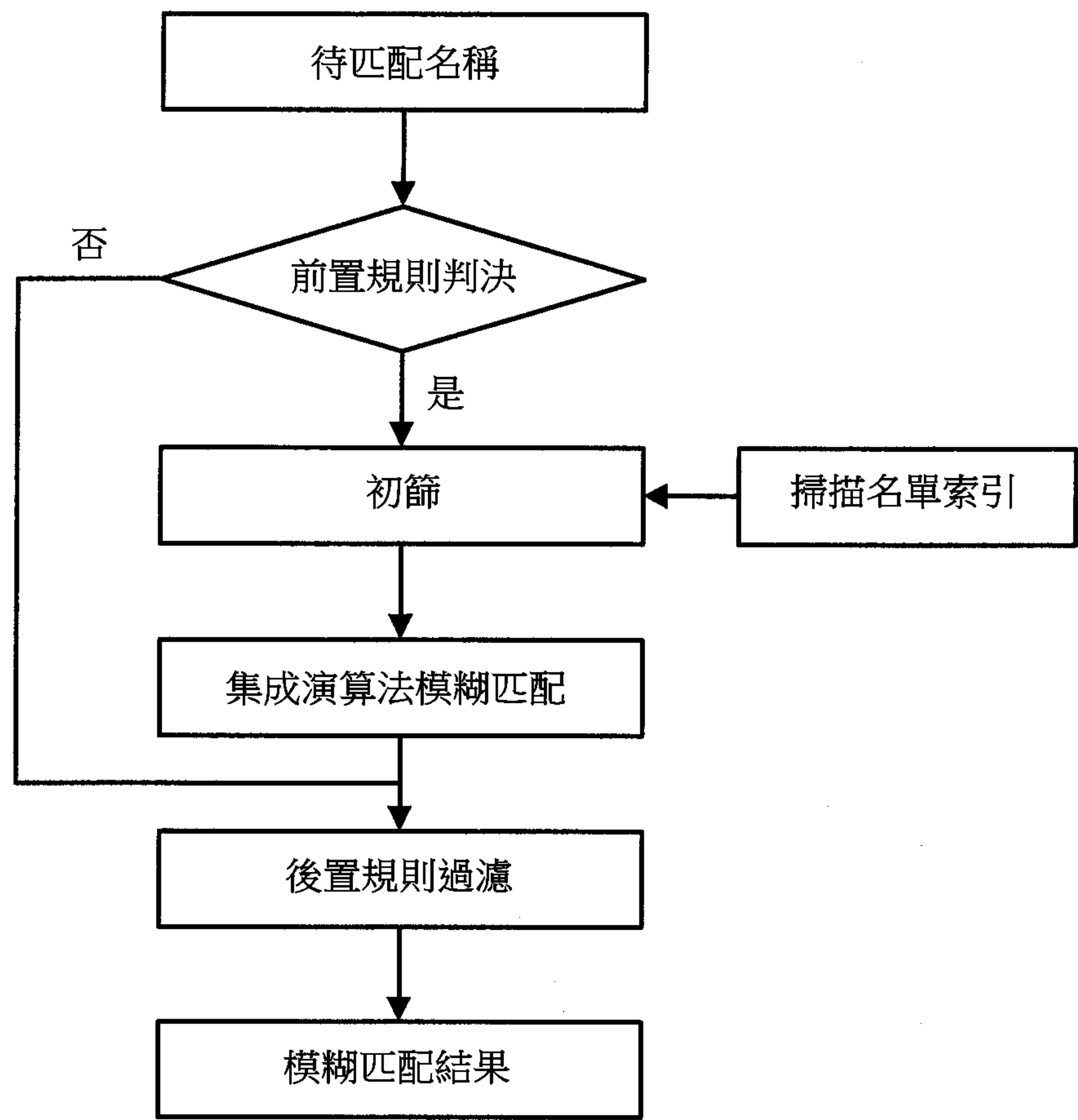


圖 4

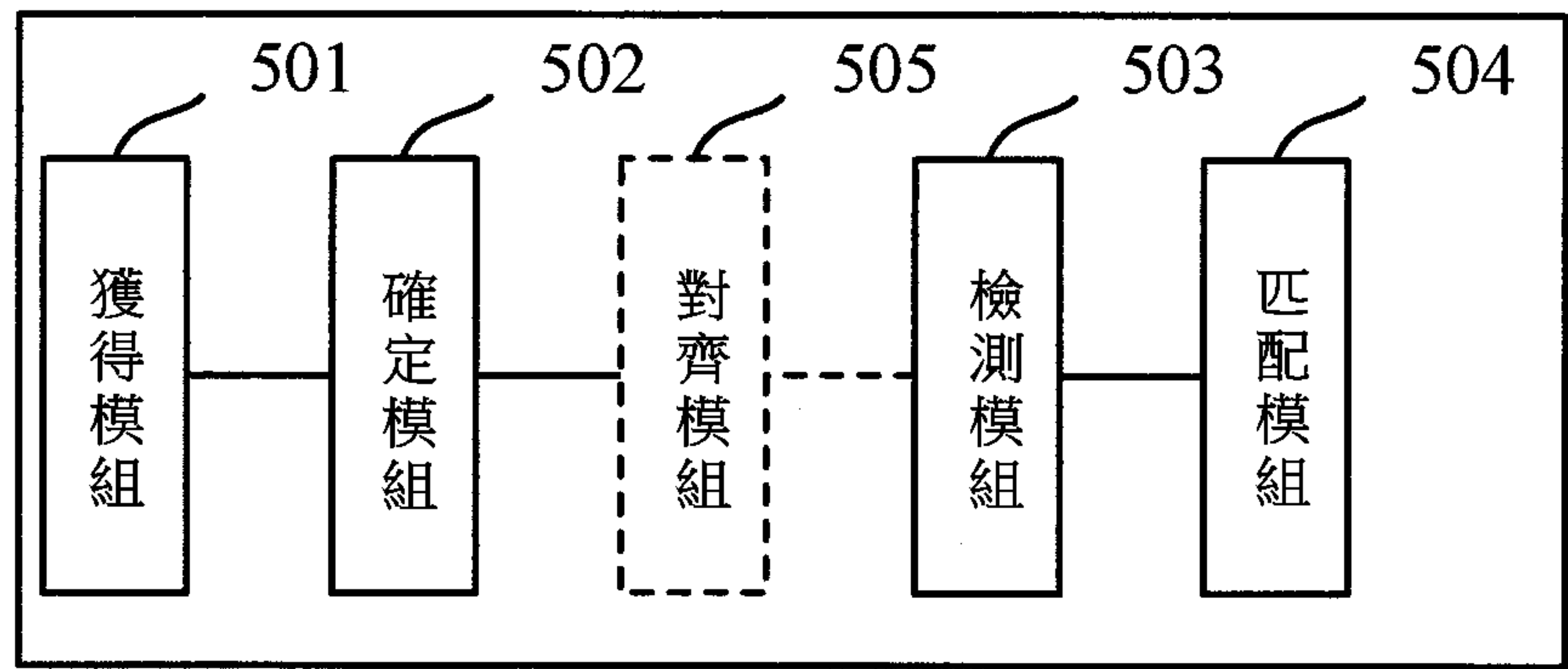


圖 5