

模糊比對分數(RC 值)說明

使用智慧型比對進行姓名檢核可提供量化的數值判斷姓名與名單資料庫的符合度高低。智慧型比對演算兩個字串（客戶姓名與資料庫姓名）之間相似的程度進行名單比對，並參考姓名規則的習慣用法（例：可判別逗號、姓名顛倒、通用字）進行調整，而演算分數的結果就是相對相似程度（Relative Correlation, RC 值分數）RC 值分數越高，則相似程度越高（符合度越高），當 RC 值分數=100 時，表示兩個字串完全相同；RC 值分數越低，表示相似程度越低（符合度越低），表示兩個字串差異越大。

*備註：其他詳細說明請參考下頁附件一。

附件一

1. 概述

全球監管單位皆要求金融機構需做客戶盡職調查(Customer Due Diligence, CDD)，其目的為了解客戶是否有涉及洗錢或資恐行為，進而要求各金融機構需持續關注客戶是否被列為國際組織公告的制裁名單。遂各金融機構便有名單過濾掃描等相關需求，但目前尚無任何監管單位提供或訂定如何去定義客戶姓名要多接近制裁名單才算是疑似洗錢或資恐名單之統一標準，於是各家名單過濾廠商便自行發展各家的獨特的演算法或邏輯進行名單比對。

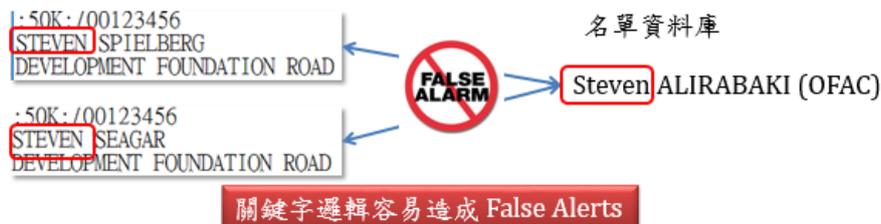
2. 基本概念

本系統 Patriot Officer 採用美國原廠 Global Vision Systems (GVS) 開發的 Global Distance Algorithm 比對方法，透過演算兩個字串(例：資料庫名單與客戶姓名)之間相似的程度進行名單比對；演算分數的結果為 Relative Correlation 值(RC 值分數)，也就是相對相似程度。

Relative Correlation 值(RC 值分數)基本概念如下：

- 若兩個字串一模一樣，代表這兩個字串是 100%相同 (RC 值分數=100)。
- 若兩個字串中有部分字母不一致或字母順序有變化，代表與 100%相似有一段距離，變化越大，代表距離越遠，所以其 RC 值分數也會越低。

模糊比對會根據相對相似度掃描資料庫名單，同時考慮姓名規則的習慣用法(請參照下一節通用規則與特殊規則的細項說明)，因此，可提供較高準確度的掃描結果。基本上關鍵字比對的原理是將輸入的關鍵字送到資料庫中去搜尋該關鍵字，所以其概念適合用在搜尋引擎，例如 Google；但不適合用在姓名比對，因為只要輸入的關鍵字在資料庫任一個名單中出現都算符合，會造成大量的錯誤警報，請參考以下案例說明：



3. 計算分數考慮因素

在演算 RC 值分數時，系統會考慮以下通用規則與特殊規則，說明如下：

3.1 通用規則

No	通用規則	比對審核之規則與說明	對 RC 值分數影響
1	字母大小寫	字母大小寫視為相同。 例如： 資料庫名單為 King Kenny，比對名字為 KING Kenny，則 RC 值分數=100。	無影響
2	符號 '(撇)	系統將忽略該符號。 例如： 資料庫名單為 John O'Connor，比對名字為 John OConnor，則 RC 值分數=100。	無影響
3	符號 \ (反斜線)	系統會視同區隔符號並忽略該符號。 例如： 資料庫名單為 King Kenny，比對名字為 King\Kenny，則 RC 值分數=100。	無影響
4	符號 () (括弧)	當有共通別名在括弧裡時，包含括弧之名字以及不包含括弧之名字將視為同一人。 例如： 資料庫名單為 John (also known as Jacky) Smith，比對名字為 John Smith，則 RC 值分數=100。	無影響
5	符號 , (逗號)	系統會視同區隔符號並忽略該符號，但會將逗號前後字串對調再掃描一次。 例如： 資料庫名單為 Smith, John，比對名字為 1. John Smith，則 RC 值分數=100。 2. Smith John，則 RC 值分數=100。	無影響
6	符號 . (點)	系統會視同區隔符號並忽略該符號。 例如： 資料庫名單為 Jake E.Black，比對名字為 Jake E Black，則 RC 值分數=100。	無影響

No	通用規則	比對審核之規則與說明	對 RC 值分數影響
7	符號 / (斜線)	系統會視同區隔符號並忽略該符號。 例如： 資料庫名單為 King/Kenny，比對名字為 King Kenny，則 RC 值分數=100。	無影響
8	多餘空白 (空白)	系統會忽略多餘空白。 例如： 資料庫名單為 King Kenny，比對名字為 King Kenny，則 RC 值分數=100。	無影響
9	其他符號	系統會視同區隔符號並忽略該符號	無影響
10	名字字數	小於等於 10 個字之名字皆可被完整比對審核	無影響
11	遺漏字母/字串	模糊比對 例如： 資料庫名單為 Fethi Ben Assen Ben Salem HADDAD，比對名字為 Fethi Ben Assen Ben HADDAD，則 RC 值分數=83。	減少 RC 值分數 (減少相似度)
12	多餘字母/字串	模糊比對 例如： 資料庫名單為 Fethi Ben Assen Ben Salem HADDAD，比對名字為 Fethi Ben Assen Ben Salem HADDAD Ali，則 RC 值分數=86。	減少 RC 值分數 (減少相似度)
13	姓名無分開	例如： 資料庫名單為 William DUMA，比對名字為 WilliamDuma，則 RC 值分數=98。	減少 RC 值分數 (減少相似度)
14	單一字母縮寫 (名字縮寫)	系統將不會另外處理單一字母縮寫之單字，因為此規則將會製造出太多不必要之可能性，因此系統無法判斷縮寫代表的全名。 例如： 資料庫名單為 Jo Ann Emerson，比對名字為 J. Ann Emerson，則 RC 值分數=83。 (因 J. 無法知道 J 代表名字為何)	減少 RC 值分數 (減少相似度)

No	通用規則	比對審核之規則與說明	對 RC 值分數影響
15	簡寫 (通用字)	所有簡寫將另外被設定在常用單詞 (Common Word List) 裡並內建通用字庫。 例如： 資料庫名單為 Williams Corporation，比對名字為 Williams Corp，則 RC 值分數=100。	無影響
16	錯誤之姓名排序	例如： 資料庫名單為 Doku UMAROV，比對名字為 UMAROV Doku，則 RC 值分數=100。	無影響
17	贅詞遺漏 (無意義字)	系統內建無意義字 例如: "Mr" 及 "Mrs" 將被視為贅詞，因此系統在辨識時會自動忽略 資料庫名單為 Osama bin laden，比對名字為 Mr. Osama bin laden，則 RC 值分數=100。	無影響
18	移除之中間名	系統將不會處理中間名，除非無中間名之名字出現在資料庫名單資料裡。 例如： 資料庫名單為 Jo Ann Emerson，比對名字為 Jo Emerson，則 RC 值分數=66。	減少 RC 值分數 (減少相似度)
19	他國翻譯之不同	他國翻譯之不同可以被設定在相等姓名表單 (Equivalent Name List) 裡。 例如： 資料庫名單為 Cheng Yun，比對名字為 Tshang Yun，將出現兩種狀況如下： 1. RC 值分數=100 當 RC 值分數結果=100 時，"Cheng" 與 "Tsang" 在相等姓名表單 (Equivalent Name List) 被設定為同一單詞。 2. RC 值分數=0 當 RC 值分數結果=0 時，表示 "Cheng" 與 "Tsang" 沒有在相等姓名表單 (Equivalent Name List) 中被設定為同一單詞，此情況將被視	無影響/ 減少 RC 值分數 (減少相似度)

No	通用規則	比對審核之規則與說明	對 RC 值分數影響
		為錯字。	
20	錯字	例如： 資料庫名單為 Osama bin laden，比對名字為 Osama bin leden，則 RC 值分數=90。	減少 RC 值分數 (減少相似度)
21	發音偵測	偵測各個單字之發音。所有單字將根據發音運算法被賦予一個條碼來辨識。 例如： 資料庫名單為 RICCIO, Mario，比對名字為 Maria Reecho，則 RC 值分數將出現以下兩種狀況： 1. 當開啟發音偵測時，RC 值分數=99 2. 當關閉發音偵測時，RC 值分數<50	根據 POCTRL 之設定來增加 RC 值分數評分
22	Non-Unicode 3 hit 2 Matches	1. 當字數對應到時將會產生結果 (不管正確或錯誤的排序) 2. 只適用於 non-unicode 3. 從名字頭或尾來比較都會產生結果 4. 如果辨識出常用單詞或相等單詞時，此規則(Non-Unicode 3 hit 2 Matches)將不符合。但是當比對出單詞且單詞並不是一個不具影響力之常用單詞，將被視為比對成功 5. 需要確認名字比對成功之單詞為具影響力之單詞(超過三個字母) 6. 單詞內有三個字母以上比對成功時，將被視為三中二。 7. 如果比對出的比例在最常的名字裡估大於 50%(i.e.. 3:4)，系統將判定為 RC 值分數 = 99。 例如：	RC 值分數 將被強制評分為 99

No	通用規則	比對審核之規則與說明	對 RC 值分數影響
		資料庫名單為 ESMAELI, Reza Gholi，比對名字為 ESMAELI, R. Gholi，則 RC 值分數=99。	
23	Space (,空白) - Special Handling 1	空白將被移除來與資料庫名單做比對審核。此功能將不會使用其他功能像是贅詞(Trivial words)、常用單詞 (common words)或相等名字 (equivalent names)。 例如： 資料庫名單為 King Kenny，比對名字為 KingKenny，則 RC 值分數=100。	無影響
24	Space (,空白) - Special Handling 2	當有一單詞完整符合時，空白將被移除來與資料庫名單做比對審核。此功能將會使用其他功能像是贅詞(Trivial words)、常用單詞(common words)或相等名字(equivalent names)。 例如： 資料庫名單為 King Kenny，比對名字為 KingKenny，則 RC 值分數=100。	無影響
25	Unicode 交錯排序比對	關閉支援中文、韓文與日文的交錯排序比對。(不會交錯所有中文、韓文與日文之單字來比對資料庫名單。) 例如： 資料庫名單為李建飛，比對名字為建飛李，不會出現比對結果。	比對結果沒有找到任何紀錄
26	檢體中文 Vs 繁體中文	檢體中文與繁體中文可以相互比對。 例如：	無影響

No	通用規則	比對審核之規則與說明	對 RC 值分數影響
		資料庫名單為刘德华，比對名字為劉德華，則 RC 值分數=100。	
27	中文四個字(含)以上者的特殊比對邏輯	字數符合 50%以上(不包括 50%)，RC 值分數即計算為 99 例如： 資料庫名單為万世集团，比對名字為万世团，則 RC 值分數=99。	RC 值分數 將被強制評分為 99

3.2 特殊規則

No	特殊規則	用途與說明	對RC值分數影響
1	常用單詞 (Common Word List)	<ol style="list-style-type: none"> 設定之常用單詞(包含簡寫以及縮寫)。 當常用單詞被設定為不太具影響力之單詞時，對於 RC 值分數之影響會較少。 將由使用者設定以及維護此功能。 此工具表單只能寫入單詞，不能寫入符號。 Unicode 字型: 不支援 可支援功能：Name Check(CDC)、FLRS、Outsource、SWIFT、RTGS、FEDWIRE <p>例如： 資料庫名單為 BOC Mutual Fund Management Co., Ltd，若比對名字為 BOC Mutual Fund Management Ltd，則 RC 值分數=90；若比對名字為 Mutual Fund Management Co Ltd，則 RC 值分數=80。</p> <p>因前者比對名字之漏字"Co"為常用單詞，因此它對於 RC 分數之影響相較於第二項漏詞的"BOC"會來的少。</p>	<p>兩種影響</p> <ol style="list-style-type: none"> 當有單詞符合常用單詞時，增加 RC 值分數(增加相似度) 當有單詞符合不太具影響力之常用單詞時，減少 RC 值分數(減少相似度)
2	贅詞表單 (Trivial Word List)	<ol style="list-style-type: none"> 系統將依據贅詞表單排除特定單字。 此功能是被系統設定裡寫入之文字及關鍵字來控制。 Unicode 字型: 不支援 可支援功能：Name Check(CDC)、FLRS、Outsource、SWIFT、RTGS、 	

No	特殊規則	用途與說明	對RC值分數影響
		<p>FEDWIRE</p> <p>例如： 資料庫名單為 Antonio el Toldo，比對名字為 Antonio Toldo，則 RC 值分數=100。因單詞 'el'被偵測到為贅詞所以排除。</p>	
3	相等姓名表單 (Equivalent Name List)	<ol style="list-style-type: none"> 設定之相等姓名將被視為暱稱。 將由使用者設定以及維護此功能。 此工具表單只能寫入單詞，不能寫入符號。 Unicode 字型：支援 可支援功能：Name Check(CDC)、FLRS、Outsource、SWIFT、RTGS、FEDWIRE <p>例如： 資料庫名單為 William Jackson，比對名字為 Will Jackson，則 RC 值分數=100。因單詞'William' 與 'Will'屬於相等姓名。</p>	
4	減少 RC 值分數之設定	<ol style="list-style-type: none"> 對於特定情況下減少 RC 值之分數，由使用者設定以及維護此功能。 Unicode 字型：不支援 可支援功能：Name Check(CDC) 	三種會降低 RC 值分數之因素 (減少相似度) <ol style="list-style-type: none"> 國家(名字對應但於不同國家時，減少 RC 值分數) 出生年月日(名字對應但出生年月日不同時減少 RC 值分數) 低風險名單(名字對應但出現於低風險名單時減少 RC 值分數)
5	白名單	<ol style="list-style-type: none"> 白名單中的名字將不會被檢查比對，只用於減少錯誤查核之情況。 將由使用者設定以及維護此功能。 建議之姓名請參考錯誤比對報告(False Hit report)。BIC 編碼無法由此報告查出 	

No	特殊規則	用途與說明	對RC值分數影響
		<p>4. 當出現一些標點符號如逗點或句點時，將視為不同名字。例：白名單的名字為 "Commercial bank co. ltd."，當出現名字為"Commercial bank co ltd"時，仍會被拿來查核比對。</p> <p>5. Unicode 字型：不支援</p> <p>6. 可支援功能： FLRS、Outsource、SWIFT、RTGS、FEDWIRE</p> <p>例如： 資料庫名單為 Commercial Bank，比對名字為 Commercial Bank of China，則 RC 值分數=0。 因當白名單裡有 Commercial Bank of China 的情況時，此資料庫名單 "Commercial Bank" 將不會拿來做比對</p>	
6	勿掃描名單 (Do-Not-Scan List)	<p>1. 當名字被查核、審核以及視為錯誤查核的情況時，此名字或相似名字在新的資料庫名單出現之前將不會被拿來比對。</p> <p>2. 將由使用者設定以及維護此功能。</p> <p>3. Unicode 字型：不支援</p> <p>4. 可支援功能： FLRS、Outsource、SWIFT、RTGS、FEDWIRE</p>	
7	極短姓名排除	<p>1. 此功能是被系統設定裡寫入之文字及關鍵字來控制。</p> <p>2. Unicode 字型：支援</p> <p>3. 可支援功能： Unicode、FLRS、Outsource、SWIFT、RTGS、FEDWIRE</p> <p>例如： 如果設定為"SHORTNAME_LEN = 3"，名字 "FED" 將會被排除審核</p>	排除
8	常見別名排除	<p>1. OFAC 有提及對於比對較常見的別名時之出錯狀況，因此系統將會排除比對較常見的別名。預設設定"ALT_STR = 7"來比對所有類型的名字。</p> <p>2. 此功能是被系統設定裡寫入之文字及關鍵字來控制。</p> <p>3. Unicode 字型：支援</p>	排除

No	特殊規則	用途與說明	對RC值分數影響
		4. 可支援功能： FLRS、Outsource、SWIFT、RTGS、FEDWIRE	
9	BIC Code Scan	<ol style="list-style-type: none"> 1. 字數在 8 跟 11 個字時會拿來與 OFAC SDN, OFAC NSDN, UN 跟 Section 311 BIC code 比對。預設設定在前 X 個字比對到時就算是比對成功: OFAC SDN = 6, OFAC NSDN = 6, UN & 311 = 4。 2. 此功能是被系統設定裡寫入之文字及關鍵字來控制。 3. Unicode 字型：不支援 4. 可支援功能： FLRS、Outsource、SWIFT、RTGS、FEDWIRE 	RC 值分數將被強制評分為 100
10	BIC Country Code	<ol style="list-style-type: none"> 1. 從 BIC 編碼取得國家編碼(position 5,6)來對於限制國家名單之國家編碼進行比對。 2. 此功能是被系統設定裡寫入之文字及關鍵字來控制。 3. Unicode 字型：不支援 4. 可支援功能： FLRS、Outsource、SWIFT、RTGS、FEDWIRE 	RC 值分數將被強制評分為 100
11	Chinese Commercial Code (CCC)	<ol style="list-style-type: none"> 1. 在特定檢核名單像是內部名單 (internal list)、全球內部名單(global internal list)以及 HKSFC 內的中文名字完全比對以及交錯排序比對。 2. 此功能是被系統設定裡寫入之文字及關鍵字來控制。 3. Unicode 字型：支援 4. 可支援功能： FLRS、Outsource、SWIFT、RTGS、FEDWIRE 	RC 值分數將被強制評分為 100