



(21)申請案號：106145557

(22)申請日：中華民國 106 (2017) 年 12 月 25 日

(51)Int. Cl. : **G06F17/00 (2019.01)**

(30)優先權：2016/12/23 美國 62/438,997

2017/01/19 美國 15/409,990

(71)申請人：美商卡隼翁有限責任公司(美國) CADREON LLC (US)

美國

(72)發明人：帕特爾 布潘達墨漢拉 PATEL, BHUPENDRA MOHANLAL (US)；帕特爾 圖沙

爾 PATEL, TUSHAR (US)

(74)代理人：陳昭明

(56)參考文獻：

TW 201520890A US 5408654

US 6094647 US 6356897B1

US 2010/0114952A1

審查人員：李榮祥

申請專利範圍項數：20 項 圖式數：18 共 61 頁

(54)名稱

電腦化方法及使用元合併及/或元分組的索引的電腦化系統

(57)摘要

本發明之實施例藉由除典型地使用的彼等臨時搜索檔案之外產生促進資料集之處理的分離資料結構或組織來改良習知或常規技術。在一個實施例中，替代使用經分類臨時檔案或運行分類演算法，元合併及/或元分組的索引可經產生或創建來對前述資料集中之記錄提供快速存取。此外，本發明之實施例消除在對前述資料集進行查詢之前或期間對分類或預分類資料集之需要。本發明之態樣使用例如前述元合併及/或元分組的索引來預組織前述資料，以使得當針對前述資料集執行查詢時，針對前述元合併及/或元分組的索引來執行前述查詢。

Embodiments of the invention improve over conventional or routine technologies by generating a separate data structure or organization, other than those temporary search files that are typically used, that facilitates the processing of the datasets. In one embodiment, instead of using sorted temp files or running sorting algorithms, a meta-join and/or meta-group-by indexes may be generated or created that provide a quick access to the records in the datasets. Moreover, embodiments of the invention eliminate the need to sort or pre-sort datasets before or during a query is conducted on the dataset. Aspects of the invention use, for example, the meta-join and/or meta-group-by indexes to pre-organize the data such that, when executing a query against the dataset, the query is executed against the meta-join and/or meta-group-by indexes.

指定代表圖：

符號簡單說明：

602 . . . 方塊

604 . . . 方塊

606 . . . 方塊

608 . . . 方塊

610 . . . 方塊

612 . . . 方塊

614 . . . 方塊

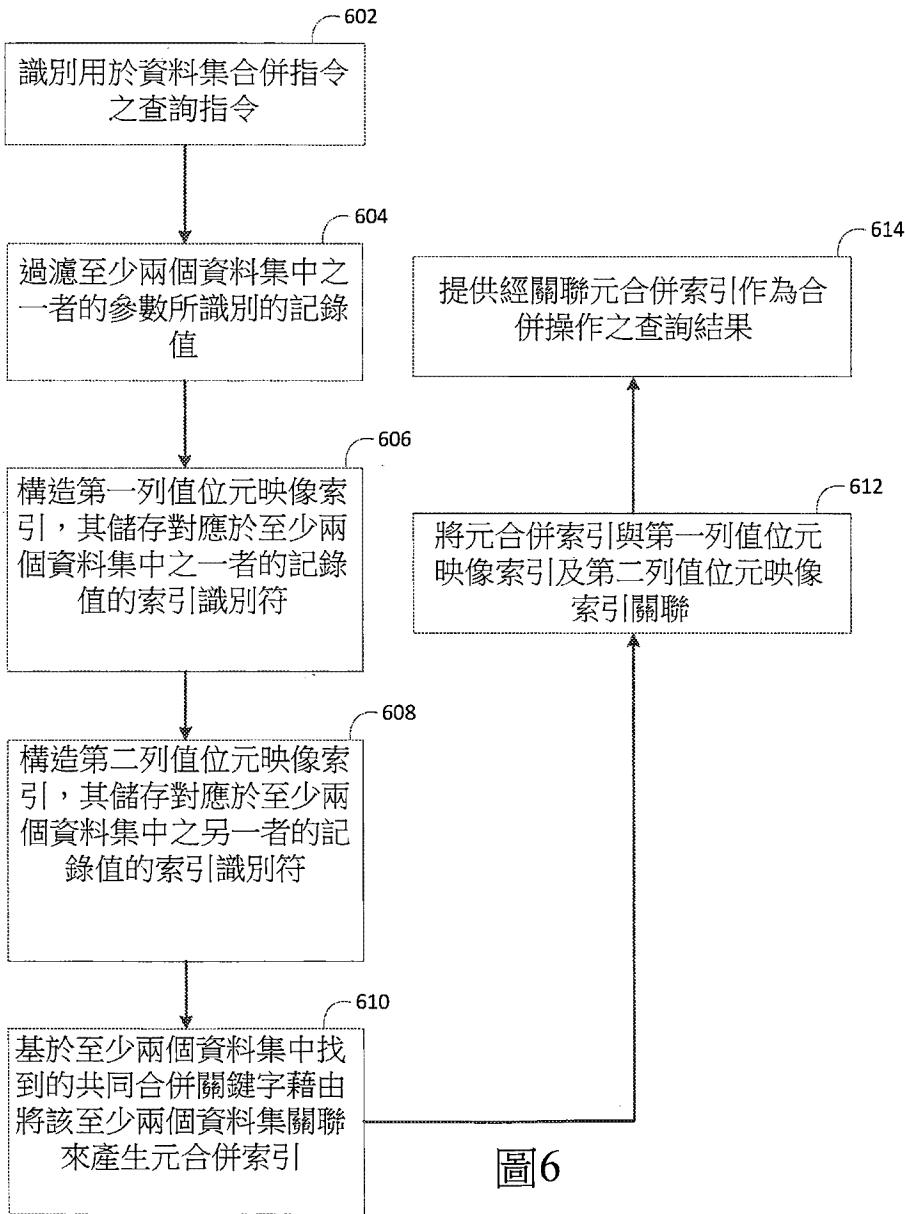


圖6

發明專利說明書

【發明名稱】

電腦化方法及使用元合併及/或元分組的索引的電腦化系統

COMPUTERIZED METHODS AND COMPUTERIZED SYSTEM USING META-JOIN AND/OR META-GROUP-BY INDEXES

【技術領域】

相關申請案之交互參照

[0001]

此為 2016 年 12 月 23 日申請的臨時申請案序列號第 62/438,997 號之非臨時申請案。上文提及的臨時申請案之整體揭示內容係以引用方式併入本文。

【先前技術】

[0002]

在資料溢流及資料超載之時代，自裝置產生、接收、處理等等的原始資料之使用者恆定地使用計算裝置來分析此等資料以找出其中的含義或識別以前已被忽略的含義。隨著雲端儲存及計算之巨大增長，資料處理或主機代管提供者持續為使用者增加資料儲存容量。同樣，隨著處理器或微處理器、以及網際網路存取速度之增加的處理能力，基於用戶端的資料處理與基於雲端的資料處理之間的差距已顯著地減小。

[0003]

對資料儲存及計算能力之恆定增加的關注似乎尤其係針對已負面地影響表格記錄資料組織結構方案及資料結構軟體程式化的問題。此問題係關於隨資料組織及/或結構變化的時間量，其旨在獲得由資料集之查詢產生的所欲資料。計算能力及資料儲存技術(例如，自硬磁碟驅動機(hard drive disk; HDD)至固體狀態驅動機(solid state drive; SSD))之增加試圖減少或減輕資料之搜索、存取、及處理的影響。然而，尤其當需要處理的資料集包括極大集合，諸如具有數百萬或數十億之記錄的資料集時，時間因數係更為顯著的。

【發明內容】

[0004]

本發明之實施例藉由產生除典型地使用來促進資料集之處理的彼等暫時性搜索檔案外的分離資料結構或組織而改良習知或常規技術。在一個實施例中，替代使用分類臨時檔案或運行分類演算法，元合併及/或元分組的索引可產生或創建來提供對資料集中之記錄的快速存取。此外，本發明之實施例消除在對資料集實施查詢之前或期間對分類或預分類資料集的需要。本發明之態樣使用例如元合併及/或元分組的索引來預組織資料，以使得當針對資料集執行查詢時，該查詢係針對元合併及/或元分組的索引來執行。

【圖式簡單說明】

[0005]

當結合隨附圖式考慮時，本發明可藉由參考詳細描述來更好地理解。圖式中之組件可能未必按比例繪製，而替代地重點是關注對本發明之原理的例示。在圖式中，相同參考符號在所有不同視圖中指定相應部件。

[0006]

圖 1 例示根據本發明之一個實施例的示範性表格，其展示有關於電視節目之消費者觀看的樣本資料集。

[0007]

圖 2 例示根據本發明之一個實施例的示範性表格，其展示有關於電視觀眾的樣本資料集。

[0008]

圖 3 例示根據本發明之一個實施例的用於使用圖 1 及圖 2 之局部唯一識別符(locally unique identifier; LUID)關鍵字合併的元合併索引之示範性表格。

[0009]

圖 4 例示根據本發明之一個實施例的用於圖 1 中之樣本資料集的列值位元映像索引之示範性表格。

[0010]

圖 5 例示根據本發明之一個實施例的用於圖 2 中之樣本資料集的列值位元映像索引之示範性表格。

[0011]

圖 6 例示根據本發明之一個實施例的流程圖，其展示用於產生元合併索引之方法。

[0012]

圖 7 例示根據本發明之一個實施例的示範性表格，其展示更新的元合併索引。

[0013]

圖 8 例示根據本發明之一個實施例的圖表，其展示示範性計算裝置。

[0014]

圖 9 例示根據本發明之一個實施例的示範性表格，其展示用於對圖 1 及圖 2 之樣本資料集進行樣本查詢的元分組的索引，該等樣本資料集係藉由 Network & Series 分組且使用 LUID 關鍵字合併。

[0015]

圖 10 例示根據本發明之一個實施例的示範性表格，其展示相異配對的資料集。

[0016]

圖 11 例示根據本發明之一個實施例的示範性表格，其展示具有值 Discovery-Life (1,4,7,8,11,12,13,15,17) 的所有記錄且將此與針對每一合併關鍵字之 TV 觀眾合併關鍵位置交叉的位元映像位置。

[0017]

圖 12 例示根據本發明之一個實施例的示範性表格，其展示具有值 ABC-Twisted (2,6,21,22,25) 的所有記錄之位元映像位置及與針對每一合併關鍵字之 TV 觀眾合併關鍵位置的交集。

[0018]

圖 13 例示根據本發明之一個實施例的示範性表格，其展示具有兩個分組資料集值的元分組的索引。

[0019]

圖 14 例示根據本發明之一個實施例的示範性表格，其展示經過濾記錄清單。

[0020]

圖 15 例示根據本發明之一個實施例的示範性表格，其展示基於圖 14 的經過濾記錄清單。

[0021]

圖 16 例示根據本發明之一個實施例的流程圖，其展示用於產生元分組的索引之方法。

[0022]

圖 17 例示根據本發明之一個實施例的用於元合併索引之示範性資料結構。

[0023]

圖 18 例示根據本發明之一個實施例的用於元分組的索引之示範性資料結構。

[0024]

一般技藝人士可瞭解的是，圖式中之要素係出於簡單及清晰度起見例示，因此並非已展示所有連接及選項以避免模糊發明態樣。例如，在商業可行實施例中 useful 或必要的一般但熟知的要素可能常常未做描繪以便有助於本揭示內容之此等各種實施例之較少妨礙視圖。將進一步瞭解，某些動作及/或步驟可以特定發生順序來描述或描繪，而熟

習此項技術者將理解，此種相對於序列而言的特定性並非實際所需的。亦將理解，本文使用的術語及表述可相對於其在調查及研究中之相應各別領域來定義，其中在本文已另外闡述特定含義的情況除外。

【實施方式】

[0025]

現可參考形成本發明之一部分且以說明方式展示本發明可藉以實踐的特定示範性實施例的隨附圖式更完全地描述本發明。可在理解本揭示內容為一或多個發明之原理之例證且可不意欲將發明任一者限制至所說明實施例的情況下來提供此等說明及示範性實施例。本發明可以許多不同的形式體現且不應解釋為限於本文闡述的實施例；實情為，此等實施例係提供來使得本揭示內容將為徹底及完整的，且將為熟習此項技術者完全地傳達本發明之範疇。尤其，本發明可體現為方法、系統、電腦可讀媒體、設備、或裝置。因此，本發明可採取完全硬體實施例、完全軟體實施例、或組合軟體及硬體態樣之實施例的形式。以下詳細描述可因此不以限制意義來進行。

[0026]

應理解，在科學領域、社會科學研究、製造等等中，所有學科的分析員恆定地面臨過多資料要識別及處理。此等分析之部分將產生來自不同資料集的見解。存在許多方法，但方法之一是識別資料集之間的相關性。在此嘗試中，可執行儲存在資料庫儲存器或結構中之兩個或兩個以

上資料集之合併操作。「合併(join)」操作可為資料庫程式化之部分且此種操作典型地包括語法及一組用於諸如資料庫伺服器的資料庫定向電腦的所需參數以執行並產生結果。當資料集較小時，合併操作不會太久，且典型地即刻提供或產生結果。然而，當資料集之大小開始指數增長時，合併效能顯著受影響。此係因為合併操作需要在產生具有合併結果的輸出資料集之前分類並搜索目標資料集中之相關資料欄位。

[0027]

作為說明而非作為限制，考慮圖 1 中展示的實例，其為例示有關於 TV 節目之消費者觀看的資料集之樣本的表，且考慮圖 2 中特定 TV 節目之 TV 觀眾之樣本資料集的表。應理解，其他類型的資料可在不脫離本發明之實施例之範疇及精神的情況下使用。

[0028]

為自資料集產生一些見解，例如，考慮以下查詢指令：

[0029]

SELECT PID, AGE, INCOME, NETWORK, SERIES

[0030]

FROM EXPERIAN, TV

[0031]

WHERE EXPERIAN.LUID = RV.LUID

[0032]

AND GENDER = 'F'

[0033]

AND SHOWDATE BETWEEN 01/04/2016 AND
01/06/2016

[0034]

AND (NETWORK LIKE 'D*' OR SERIES LIKE
'L*')

[0035]

AND VIEWEDFOR > 40

[0036]

作為另一假設，假設消費者觀看資料集中之記錄數目為 2 億個記錄，且 TV 觀眾資料集中之記錄數目為 20 億個記錄。

[0037]

上文查詢將藉由大多數當前資料庫解決方案，使用作為實例的以下方法之一來解決：

[0038]

巢套迴路合併(Nested Loop Join)

[0039]

散列合併(Hash Join)

[0040]

分類-歸併合併(Sort-Merge Join)

[0041]

利用引起效能問題的任何傳統合併方法之挑戰為：

[0042]

記錄尋覓時間（自硬磁碟驅動機或 SSD）。

[0043]

臨時表格產生(用於分類)

[0044]

可假定資料集可儲存在「雲端」中，但應理解，用於此種大型資料集之基於雲端的解決方案可實際上相交上文所假設者更差。基於雲端的解決方案需要資料之實體儲存，且基於雲端的解決方案實際上在使用者尋覓可實體上儲存或容納在不同位置中的資料集時具有對記錄尋覓時間及臨時表格產生之資料傳輸時間及成本。

[0045]

作為另一說明，以下實例證明尤其當在合併兩個或兩個以上表格時需要進行多次尋覓時在 HDD 中之檔案尋覓之負擔。

[0046]

假設每個檔案塊存在 100K 記錄，每百萬個記錄可能需要 10 次尋覓，每十億個記錄需要 10000 次尋覓 = $100,000 \times 2$ 毫秒 = 200 秒。若使用 SSD，則尋覓時間將減少，但其仍將耗費約 50 秒之估計值或因此僅僅進行針對十億個記錄的檔案尋覓。

[0047]

例如大約 50 秒至 200 秒的上述時間估計值僅僅用於尋覓操作以尋覓資料而甚至不需執行跨資料集的任何種類的合併操作。

[0048]

同時，藉由資料庫管理及記憶體管理之構建進行的現有實踐將不可避免地創造臨時表格以用於涉及大量記錄之合併，此係歸因於分類記錄以用於計算聚合函數之需要。臨時表格之創建、尋覓用於臨時表格之記憶體儲存、寫入及讀取臨時表格等等進一步增加總體查詢處理時間且實質上進一步減少查詢效能。

[0049]

根據本發明之一個實施例，「元合併索引(Meta-Join Index)」結構可經創建或產生以消除在運行時間不得不進行合併之需要。此方法尤其適用於查詢執行。在另一實施例中，元合併索引可提供索引識別符，例如對另一資料集之記錄的指標。

[0050]

現參考圖 3，示範性表格展示根據本發明之一個實施例的元合併索引，其利用局部唯一識別符(LUID)關鍵字來合併圖 1 及圖 2 中展示的樣本資料集。在一個實施例中，元合併索引可以表格格式提供。應理解，可在不脫離本發明之態樣之範疇或精神的情況下使用其他格式或表格尺寸。

[0051]

在圖 3 中例示的實例中，表格 300 展示合併兩個表格即圖 1 中之表格 100 及圖 2 中之表格 200 的結果。例如，在圖 1 中之表格 100 中，欄位 102 可識別與局部唯一識別

符(LUID)「L1」相關聯的#1及#2。在表格 200 中，欄位 202 可識別與相同的共同合併關鍵字 LUID L1 相關聯的 #1、#6、#10、#13、#19、及#20。在合併兩個表格時，藉由列 302 中之 LUID 組織的圖 3 中之元合併索引表格 300 可提供對相應表格的索引。例如，在列 304 中，表格 300 藉由參考數字 102 提供對表格 100 之#1及#2的索引參考(如藉由標籤「消費者觀看」所識別的)。類似地，在列 306 中，表格 300 藉由參考數字 202 提供對表格 200 之#1、#6、#10、#13、#19及#20的索引參考(如藉由標籤「TV 觀眾」所識別的)。

[0052]

在一個實施例中，元合併索引可儲存具有有效合併條件匹配的每一資料集記錄之記錄位置的關鍵字-值對。在另一實施例中，元合併索引可每個分區被分段且分區關鍵字係可組配的。

[0053]

在圖 3 中展示的此實施例中，表格 300 出於簡化目的使用且並非作為限制。在此實例中，表格 300 並非藉由分區來劃分。然而，在其他實例中，表格 300 或元合併索引可經分區，例如以日期(範圍)分區一或多次。例如，元合併索引可針對所有資料集、針對每一日期、每個月、每個季度等等基於資料分區需要來創建。

[0054]

在另一實施例中且為促進獲得查詢結果之更快方式，可使用列值位元映像索引。在一個實例中，列值位元映像索引可用來允許壓縮尺寸值至就資料記憶體或儲存器而言極小的覆蓋區。此優點可致能記憶體中之輕易加載。此外，此壓縮尺寸可基於指定在針對每一資料集之執行查詢中的 WHERE 條件而用於濾出任何記錄。在一個實例中，可藉由使用位元之自然定位以與彼列值之記錄位置匹配且消除針對記憶體中之每一位元儲存實際位置值的需要來減少對位元映像索引之儲存空間要求。在一個實例中，圖 4 例示示範性表格，其展示用於如圖 1 中所見的消費者觀看資料集之樣本的列值位元映像索引。應理解，並非所有尺寸及其列值位元映像展示於整個消費者觀看資料集之上。僅出於說明目的且並非作為限制，此處僅展示上文查詢實例之執行所需者。例如，如圖 1 所示，表格 100 指示欄位 104 指定以 #1 識別的使用者為女性，且欄位 106 指示使用者係 50 歲。在列值位元映像索引中，104 及 106 中之相同使用者之相同資訊分別反映在表格 400 中之 402 及 404 中。

[0055]

類似地，圖 5 例示根據本發明之一個實施例的用於圖 2 中之樣本資料集的列值位元映像索引之示範性表格。作為說明，表格 500 表示與分別由 502 及 504 中之 204 及 206 所表示相同的資訊。如上文所解釋，列值位元映像索引減少對表示相同資訊的記憶體儲存需要且可顯著地縮

短資料處理時間。應理解，並非所有尺寸及其列值位元映像展示於整個消費者觀看資料集之上。僅出於說明目的且並非作為限制，此處僅展示上文查詢實例之執行所需者。

[0056]

在一個實施例中，利用所構造的元合併索引及列值位元映像索引，可使用兩個資料結構來以較好的效能執行查詢且不需要高的硬體成本，使用其他傳統或常規解決方案會導致高的硬體成本。在另一實施例中，替代直接針對資料集或針對資料集之複本執行查詢，計算裝置可首先在針對資料集執行完整查詢之前自該資料集構次元合併索引及列值位元映像索引。在另一實施例中，用於任何給定資料集之元合併索引及列值位元映像索引可在計算之前或在運行時間之前自動地構造。在另一實施例中，用於任何給定資料集之元合併索引及列值位元映像索引可在接收或執行查詢或合併操作之前根據使用者指令或請求來構造。

[0057]

參考圖 6，流程圖展示根據本發明之一個實施例的用於產生元合併索引之方法。在一個實施例中，元合併索引可用來執行下文的樣本查詢。在此實例中，可使用列值位元映像索引。

[0058]

```
SELECT PID, AGE, INCOME, NETWORK, SERIES
```

[0059]

```
FROM EXPERIAN, TV
[0060]
WHERE EXPERIAN.LUID = RV.LUID
[0061]
AND GENDER = 'F'
[0062]
AND SHOWDATE BETWEEN 01/04/2016 AND
01/06/2016
[0063]
AND (NETWORK LIKE 'D*' OR SERIES LIKE
'L*')
[0064]
AND VIEWEDFOR > 40
[0065]
```

使用上文作為實例的查詢指令集，在 602，識別針對複數個資料集的查詢指令中的用於至少兩個資料集之合併操作。例如，可能存在系統或電腦接收的一組或複數個查詢指令。例如，指令可自使用者直接接收或可自自動化或經排程通道接收。例如，指令可在批次檔案中接收。在所接收指令中，例如，合併操作指令可為該等指令之一且本發明之實施例自複數個指令識別該等合併操作指令。此外，如上文所例示，合併操作指令包括參數、語法規則等等。例如，合併操作指令可包括諸如「WHERE」的參數或條件以作為合併操作之一部分。在 604，過濾合併操作

中識別至少兩個資料集中之一者的記錄值的參數。例如，在「WHERE」指令中，例如「GENDER = F」的參數識別圖 1 中之表格 100 中的記錄值。另一方面，在此實例中，「WHERE」指令未能包括參數「GENDER = M」。因此，可濾出不與參數「GENDER = F」相關聯的任何記錄值。在 606，第一列值位元映像索引可經構造來儲存對應於藉由經過濾參數識別的至少兩個資料集之一者中的記錄值的索引識別符。使用表 1 作為實例，因為「WHERE」指令需要參數為「Gender = F」，所以與表 1 相關聯的滿足此種標準的記錄值將為針對:1,2,4,5,7,9 之記錄。

[0066]

在 608，將 606 中之相同方法應用於其他資料集。例如，在 WHERE 指令中，存在另外的參數:(NETWORK LIKE ' D * ' OR SERIES LIKE ' L * ') AND VIEWEDFOR > 40，且其中每一者識別藉由參數識別的對應資料值。因此，第二列值位元映像索引可經構造來儲存對應於藉由經過濾參數識別的至少兩個資料集之另一者中的記錄值的索引識別符。當然要理解的是，過濾過程將包括首先識別參數，且隨後在可開始過濾過程之前識別資料集中藉由參數識別的資料值。此外，僅為達簡單目的且不作为限制，上文實例並非以日期來分區。若是，則步驟 602 至 608 可在配合 WHERE 條件之每一分區應用。

[0067]

在一個實例中，藉由使用針對 3 個列-**Network**、**Series** & **VIEWEDFOR** 的位元映像索引，將獲得以下記錄位置：
1,4,7,9,11,12,13,17。

[0068]

在 610，元合併索引係基於在至少兩個資料集中找到的共同合併關鍵字藉由將該至少兩個資料集關聯來構造或產生。例如，如先前圖 1 至圖 3 中所描述，元合併索引可藉由識別表格 100 及表格 200 中之每一共同合併-關鍵字來構造。在構造之後，在 612，元合併索引進一步與來自 606 之第一列值位元映像索引及來自 608 之第二列值位元映像索引關聯或交叉。

[0069]

例如，對 **LUID L1** 而言，表格 300 將列 304 (來自 102) 與來自第一列值位元映像索引之 1,2,4,5,7,9 交叉以在交叉之後得到 1,2。類似地，對於共同合併-關鍵字 **L1** 而言，表格 300 將列 306 (來自 202) 與來自第二列值位元映像索引之 1,4,7,9,11,12,13,17 交叉以得到 1,13。

[0070]

在 614，在完成合併操作指令之前提供經關聯元合併索引作為結果。例如，在一個實施例中，圖 7 例示用於根據上文所說明之步驟的經更新或經關聯元合併索引的表格 700。在此實例中，擴展需要包括在 **L1** 之結果中的所有記錄。換言之，值之置換可藉由審查兩個資料集來創建。

[0071]

1-1

[0072]

1-13

[0073]

2-1

[0074]

2-13

[0075]

應理解，可針對所有合併關鍵字執行相同過程或步驟且將基於元合併索引設定之最終結果返回至使用者。

[0076]

如將在上文實例中易於識別的，本發明之實施例藉助於元合併索引而消除或避免對用於合併操作之任何分類、臨時表格創建或進行檔案尋覓 100K 次的需要。替代直接對資料集執行合併操作指令-因此需要分類、臨時表格創建或進行檔案尋覓-本發明之態樣對元合併索引執行。藉由消除或不進行資料集之大部分的分類及大量的檔案尋覓，本發明之態樣使用極為經濟的硬體，僅僅藉由利用元合併索引連同位元映像索引而實質改良總體查詢效能。本發明之實施例克服處理合併操作指令之習知及常規方法。

[0077]

在另一實施例中，元分組的索引可經構造以進一步促進對資料集執行之操作。使用下文的一組示範性資料庫查詢指令作為實例。

[0078]

SELECT NETWORK, SERIES, COUNT(PID)

[0079]

FROM EXPERIAN, TV

[0080]

WHERE EXPERIAN.LUID = RV.LUID

[0081]

AND GENDER = 'F'

[0082]

AND VIEWEDFOR > 40

[0083]

GROUP BY NETWORK, SERIES

[0084]

當前可利用的電腦軟體產品需要進行資料之分類或混排，該資料需要「分組(group-by/group by)」子句。此要求係在具有「分組」子句的查詢執行中執行並引起顯著效能壓縮的顯著高成本的操作。因為資料集大小為大的(例如，數百萬或數十億資料值)，所以分類處理再次涉及創建臨時檔案並針對臨時表格資料集執行集合函數，且因此產生緩慢的總體查詢執行。

[0085]

在一個實施例中，「元分組的索引」消除對針對查詢執行在運行時間不得不進行分類或混排的需要。例如，元分組的索引儲存基於根據列值之相異組的每一資料集記

錄之記錄位置的關鍵字-值對。在一個實例中，元分組的索引可使用上文所述的預創建元合併索引及列值位元映像索引來計算。

[0086]

圖 9 展示用於上文樣本查詢的元分組的索引 900，其具有藉由 Network & Series 分組且使用 LUID 關鍵字合併的資料集。為獲得此種元分組的索引，圖 10 至圖 12 及圖 16 用於例示此種實施例。

[0087]

在此實例中，分組條件包括 GROUP BY NETWORK、SERIES。因此，在圖 10 中，表格 1002 例示每一 Network-Series 對。

[0088]

另外，圖 11 例示表格 1100，其用於藉由進行適當的列值位元映像索引與 TV 觀眾合併關鍵位置之交叉來找出匹配 TV 觀眾分組位置。例如，對於「Discovery-Life」對，基於圖 2，記錄為：[1,4,7,8,11,12,13,15,17]。

[0089]

圖 11 例示針對具有「Discovery-Life」之值 (1,4,7,8,11,12,13,15,17) 的所有記錄且將此與針對每一合併關鍵字的 1102 中之 L1 的 TV 觀眾合併關鍵位置交叉的位置。

[0090]

因此，對於 LUID L1 而言，在一個實施例中，將為：

[0091]

1, 4, 7, 8, 11, 12, 13, 15, 17 \cap 1, 6, 10, 13, 19, 20 產生如圖 11 中 1104 中所識別的 1 及 13。

[0092]

對於與相同「Discovery-Life」對交叉的 LUID L2 而言，交叉產生：

[0093]

1, 4, 7, 8, 11, 12, 13, 15, 17 \cap 2, 7, 11, 14, 18, 21 = 如圖 11 中 1106 中所識別的 7 及 11。

[0094]

在 TV 觀眾中之關鍵字與 TV 節目之消費者觀看之間，針對合併關鍵字中之每一者進行上文交叉。在另一實施例中，可針對不同的 NETWORK-SERIES 分組，諸如「ABC-Twisted」(亦即，「ABC」係 NETWORK 且「Twisted」係 SERIES)分組進行另一交叉，且位置為如圖 12 所示的 [2,6,21,22,25] (根據圖 2)。

[0095]

例如，對與「ABC-Twisted」對交叉的 LUID L1 而言，交叉產生：

[0096]

2, 6, 21, 22, 25 \cap 1, 6, 10, 13, 19, 20 產生如在圖 12 中之表格 1200 中的 1202 中所識別的 6。

[0097]

在另一實例中，對與「ABC-Twisted」對交叉的 LUID L2 而言，交叉產生：

[0098]

2, 6, 21, 22, 25 \cap 2, 7, 11, 14, 18, 21 產生如圖 12 中之表格 1200 中的 1204 中所識別的 2 及 21。

[0099]

一旦獲得具有每一合併關鍵字之表格，即可藉由針對每一獨特的 Network-Series 值進行消費者觀看合併關鍵位置與 TV 觀眾分組的位置之置換而計算元分組的索引。此可針對用於「Discovery」-「Life」及「ABC」-「Twisted」之分組的正好兩個 Network-Series 值產生藉由在先前部分中在圖 13 中的表格 1300 中的元分組的索引展示的表格。

[0100]

以下為上文展示的樣本查詢指令及使用元分組的索引執行此查詢的步驟序列之另一實例。

[0101]

```
SELECT NETWORK, SERIES, COUNT(PID)
```

[0102]

```
FROM EXPERIAN, TV
```

[0103]

```
WHERE EXPERIAN.LUID = RV.LUID
```

[0104]

```
AND GENDER = 'F'
```

[0105]

AND VIEWEDFOR > 40

[0106]

GROUP BY NETWORK, SERIES

[0107]

使用圖 16 作為實例來說明示範性方法，在 1602，接收用於分組至少兩個資料集中之資料的資料庫查詢指令。例如，存在系統或電腦接收的一組或複數個查詢指令。例如，指令可自使用者直接接收或可自自動化或經排程通道接收。例如，指令可在批次檔案中接收。在所接收指令中，例如，合併操作指令可為該等指令之一且本發明之實施例自複數個指令識別該等合併操作指令。此外，如上文所例示，合併操作指令包括參數、語法規則等等。

[0108]

在 1604，識別所接收資料庫查詢指令中用於分組至少兩個資料集中之資料值的分組條件。另外，在 1606 識別合併操作指令。在一個實例中，「WHERE」語句或條件可為合併操作指令之實例。因此，在 1608，構造第一列值位元映像索引，其儲存對應於至少兩個資料集之一者中的記錄值的索引識別符。例如，合併操作指令可包括諸如「WHERE」的參數或條件，其指示「Gender = F」作為標準之一，因此吾等將需要具有根據圖 1 中之表格 100 的消費者觀看資料集中之以下位置的記錄。例如，使用圖

1 中之表格 100，可過濾的位置可為：#1、#2、#4、#5、#7、及#9。

[0109]

類似地，回應於過濾 TV 觀眾資料集之所有記錄位置且僅保持需要使用列值位元映像索引之記錄位置，在 1610 構造第一列值位元映像索引，其儲存對應於至少兩個資料集之一者中的記錄值的索引識別符。在上文實例中，使用「Discovery-Life」之 NETWORK-SERIES 實例作為實例，位置為：#1、#2、#4、#6、#7、#9、#10、#11、#12、#13、#16、#17、#18、#19、#23、及#24(基於圖 2 中之表格 200 中之資料值)。

[0110]

在一個實施例中，針對元分組的索引中之 Network-Series 值中之每一者，應用過濾器來僅包括具有上文位置之記錄。

[0111]

此外，一旦應用合併操作指令之「WHERE」部分中之「VIEWEDFOR > 40」條件且滿足「分組」條件，即可在 1614 構造圖 15 中之元分組的索引表格 1500 從而展示分組結果。在一個實施例中，可在期中使用圖 6 及圖 7 中之圖解構造元合併索引。例如，使用圖 14 中之表格 1400 作為實例，在 1612 構造物合併索引，其展示上文關於消費者觀看資料集及「Discover-Life」之 Network-Series 對的示範性位置。

[0112]

在另一實施例中，在 1616，可針對每一 Network-Series 值之經過濾記錄清單中之每一者執行聚合函數。在圖 15 中例示的實例中，聚合函數可展示為「9」的「Discover-Life」對之結果。

[0113]

如將自 1602 至 1616 所瞭解的，基於本發明之實施例，不需要進行任何分類、臨時表格創建或進行檔案尋覓 100K 次。藉由消除分類資料集之大部分及大量檔案尋覓之問題，本發明之實施例使用極為經濟的硬體，僅僅藉由利用元分組的索引連同元合併索引及如上文所示針對資料集預計算的位元映像索引而實質改良總體查詢效能。

[0114]

如同提供於本揭示內容中之其他實例一樣，出於簡單性來給出實例且並非作為限制，僅兩個列用於分組，上文邏輯適用於可被包括在分組中的任何數量之列。

[0115]

圖 17 例示根據本發明之一個實施例的資料結構 1702，其具有用於儲存元合併索引之資料欄位。例如，如上文關於圖 3 及圖 7 所述，資料結構 1702 可包括用於共同合併-關鍵字之欄位 1704。例如，使用圖 7 作為實例，共同合併-關鍵字為 LUID 列 702。因此，第一列中之欄位儲存 LUID 值。此外，資料結構 1702 包括一或多個欄位 1706-1 至 1706-n (其中 $n > 0$)，每一者儲存用於藉由共同

合併-關鍵字識別的資料值之索引值。再次，使用圖 7 作為實例，兩個列 704 及 706 儲存在兩個合併資料集中識別的索引值：消費者觀看及 TV 觀眾。應理解，可在不脫離本發明之範疇或精神的情況下，其他數量之資料集可由資料結構 1702 表示。

[0116]

作為另一說明，圖 18 提供根據本發明之一個實施例的圖表，其展示資料結構 1802，具有用於儲存元分組的索引之資料欄位。使用圖 15 作為實例，資料結構 1802 包括用於識別分組條件之欄位 1804。例如，列 1502 包括用於識別特定分組條件之欄位，該特定分組條件諸如「Discovery-Life」之「Network-Series」對。應理解，可在不脫離本發明之實施例之範疇或精神的情況下使用其他分組條件。

[0117]

此外，資料結構 1802 包括另外的資料欄位用於 1806-1 至 1806-n (其中 $n > 0$)，每一者儲存用於作為合併及分組條件之結果而識別的資料值之索引值。再次，使用圖 15 作為實例，兩個列 1504 及 1506 儲存在兩個合併資料集中識別的索引值。應理解，資料欄位 1806 及 1706 不儲存實際資料值；其在資料集中儲存識別有關資料值之索引值。如所解釋的，其對於執行合併或分組操作為有益的及經濟的，尤其在大型資料集中如此。

[0118]

進一步應理解，呈計算裝置或電腦形式的電腦系統可用以執行上文說明的電腦可執行指令。此種電腦系統可在圖 8 例示為計算裝置 841。在高位準下，計算裝置 841 可包括數位儲存器，諸如磁碟、光碟、快閃儲存器、非依電性儲存器等等。結構化資料可儲存在諸如資料庫之數位儲存器。計算裝置 841 可具有處理器 1000，其根據電腦可執行指令來實體上組配。亦可具有聲音及視訊模組 1005，其輔助顯示視訊及聲音，且可在不使用時關閉以保存電力及電池壽命。計算裝置 841 亦可具有依電性記憶體 1010 及非依電性記憶體 1015。

[0119]

資料庫 1025 可儲存在記憶體 1010 或 1015 或可為分離的。資料庫 1025 亦可為計算裝置 841 之雲端之部分，且可以橫跨複數個計算裝置 841 的分配方式儲存。亦可存在輸入/輸出匯流排 1020，其往返於各種使用者輸入裝置運送資料，該等使用者輸入裝置諸如麥克風、攝影機、諸如輸入墊、顯示器、及揚聲器等等之輸入端。輸入/輸出匯流排 1020 亦可控制經由無線或有線裝置與網路的通訊。在一些實施例中，應用可在局部計算裝置上且在其他實施例中，應用可為計算裝置 841。當然，此僅為計算裝置 841 之一個實施例且可攜式計算裝置 841 之數量及類型僅受想像力之限制。

[0120]

所主張系統及方法可解決若干技術問題及挑戰，其中一些已描述。當前，跨於網路輸入潛在的敏感資料使得使用者緊張程度達到可能損失銷售額或可能接收不到金錢或時間節省的小費或贈券之點。藉由使用諸如支付網路之專有網路來傳輸潛在敏感資料，安全性可更高且使用者可更開放以合併另外的有益的程式。類似地，將資料自一個支付系統移動至另一忠誠系統對一些使用者來說感受到危險，但藉由使用專有、可信賴網路，資料可以更值得信賴之方式傳達。另外，以可藉由各種另外程式理解的方式格式化資料及傳達資料係系統及已解決的技術挑戰或問題。

[0121]

本文描述的使用者裝置、電腦及伺服器可為通用電腦，其可具有微處理器(諸如來自 Intel Corporation、AMD 或 Motorola)；依電性及非依電性記憶體；一或多個大量儲存裝置(亦即，硬碟驅動機)；各種使用者輸入裝置，諸如滑鼠、鍵盤、或麥克風；及視訊顯示系統以及其他元件。本文描述的使用者裝置、電腦及伺服器可在許多作業系統中之任一者上運行，該等作業系統包括但不限於 WINDOWS、UNIX、LINUX、MAC OS、或 Windows (XP、VISTA 等等)。然而，涵蓋的是，任何適當作業系統可用於本發明。伺服器可為網頁伺服器之叢集，該等網頁伺服器可各自為基於 LINUX 的且藉由負載平衡器支援，該負

載平衡器決定網頁伺服器之該叢集中之何者應基於可利用伺服器之當前請求-負載來處理請求。

[0122]

本文描述的使用者裝置、電腦及伺服器可經由網路通訊，該等網路包括網際網路、WAN、LAN、Wi-Fi、其他電腦網路(現在已知的或將來發明的)、及/或前述者之任何組合。藉由獲得本說明書、圖式、及申請專利範圍之一般技藝人士應理解的，網路可經由有線及無線導管(包括銅)、光纖、微波、及其他形式之無線電頻率、電學及/或光學通訊技術之任何組合來連接各種組件。亦應理解，任何網路可以不同方式連接至任何其他網路。系統中電腦與伺服器之間的互連為實例。本文描述的任何裝置可經由一或多個網路與任何其他裝置通訊。

[0123]

示例性實施例可包括除所示之彼等者外的另外裝置及網路。另外，描述為藉由一個裝置執行之功能性可藉由兩個或兩個以上裝置分配並執行。多個裝置亦可組合成單一裝置，其可執行組合裝置之功能性。

[0124]

本文描述之各種參與者及元件可操作一或多個電腦設備以促進本文描述的功能。包括任何伺服器、使用者裝置、或資料庫的上述圖式中的元件中之任何者可使用任何適當數量之子系統來促進本文描述的功能。

[0125]

本申請案中描述的軟體組件或功能中之任何者可作為軟體代碼或電腦可讀指令來實施，該等指令可藉由至少一個處理器，使用任何適當的電腦語言，使用例如習知或物件導向技術來執行，該等電腦語言諸如例如 Java、C++、或 Perl。

[0126]

軟體代碼可儲存為非暫時性電腦可讀媒體上之一系列指令或命令，該非暫時性電腦可讀媒體諸如隨機存取記憶體(random access memory; RAM)、唯讀記憶體(read only memory; ROM)、磁性媒體諸如硬碟機或軟碟、或光學媒體諸如 CD-ROM。任何此種電腦可讀媒體可駐留於單一計算設備上或內且可在系統或網路內之不同計算設備上或內存在。

[0127]

可以理解的是，如上文所述的本發明可以控制邏輯形式使用電腦軟體以模組或整合方式實施。基於本文提供的揭示內容及教示內容，一般技藝人士可知曉且瞭解使用硬體、軟體、或硬體及軟體中之組合實施本發明的其他方式及/或方法。

[0128]

上文描述係說明性的且不為限制性的。在回顧本揭示內容之後，本發明之許多變化將對熟習此項技術者而言顯而易見。因此，本發明之範疇應不參考上文描述而決定，

但替代地應參考待決申請專利範圍連同其完整範疇或等效物決定。

[0129]

來自任何實施例之一或多個特徵可與任何其他實施例之一或多個特徵組合而不脫離本發明之範疇。除非明確地指示為相反，否則「一」或「該」之敘述意欲指「一或多個」。除非明確地指示為相反，否則「及/或」之敘述意欲表示術語之最具包括性的意義。

[0130]

本發明系統之元件中之一或多個可作為用於完成特定功能之手段來主張。在此種手段-加-功能元件用於描述所主張系統之某些元件時，獲得本說明書、圖式及申請專利範圍之一般技藝人士將理解的是，對應結構係經程式化以使用在未特殊程式化的任何通用電腦中找到的功能性及/或藉由實施一或多種演算法來達成所敘述功能性來執行特定敘述功能的通用電腦、處理器、或微處理器(視情況而定)。如一般技藝人士將理解的，演算法可在本揭示內容內表示為數學公式、流程圖、敘說、及/或呈為一般技藝人士提供充分結構來實施所敘述過程及其等效物之任何其他方式。

[0131]

雖然本揭示內容可以許多不同的表格體現，但圖式及論述係在理解本揭示內容為一或多個發明之原理之例證

的情況下提出且不意欲將本發明之任一者限制於所說明的實施例。

[0132]

本揭示內容提供對上文所述的長期需要之解決方案。詳言之，本文描述的系統及方法可經組配以改良資料酬載執行系統。上文描述的系統及方法之其他優點及修改將易於由熟習此項技術者思及。因此在本揭示內容之較廣態樣中，本揭示內容不限於特定細節、代表性系統及方法、及上文所示及所述的說明性實例。在不脫離本揭示內容之範疇或精神的情況下，可對上文說明書做出各種修改及變化，且本揭示內容意欲覆蓋所有的此種修改及變化，前提是該等修改及變化落入以下申請專利範圍及其等效物之範疇內。

【符號說明】

[0133]

100	表格
102	欄位
104	欄位
106	欄位
200	表格
202	欄位
204	欄位
206	欄位
300	表格

302	列
304	列
306	列
400	表格
402	列
404	列
500	表格
502	列
504	列
602	方塊
604	方塊
606	方塊
608	方塊
610	方塊
612	方塊
604	方塊
700	表格
702	列
704	列
706	列
841	計算裝置
900	表格
1000	處理器
1002	表格

1005	聲音及視訊模組
1010	依電性記憶體
1015	非依電性記憶體
1020	輸入/輸出匯流排
1025	資料庫
1100	表格
1102	欄位
1104	欄位
1106	欄位
1200	表格
1202	欄位
1204	欄位
1300	表格
1400	表格
1500	表格
1502	列
1504	列
1506	列
1602	方塊
1604	方塊
1606	方塊
1608	方塊
1610	方塊
1612	方塊

- 1614 方塊
- 1616 方塊
- 1702 資料結構
- 1704 欄位
- 1706-1 資料欄位
- 1706-2 資料欄位
- 1706-3 資料欄位
- 1802 資料結構
- 1804 欄位
- 1806-1 資料欄位
- 1806-2 資料欄位
- 1806-3 資料欄位

I677795

發明摘要

※ 申請案號：106145557

※ 申請日：106年12月25日 ※IPC 分類：G06F 17/30 (2006.01)

【發明名稱】

電腦化方法及使用元合併及/或元分組的索引的電腦化系統

COMPUTERIZED METHODS AND COMPUTERIZED SYSTEM USING META-JOIN AND/OR META-GROUP-BY INDEXES

【中文】

本發明之實施例藉由除典型地使用的彼等臨時搜索檔案之外產生促進資料集之處理的分離資料結構或組織來改良習知或常規技術。在一個實施例中，替代使用經分類臨時檔案或運行分類演算法，元合併及/或元分組的索引可經產生或創建來對前述資料集中之記錄提供快速存取。此外，本發明之實施例消除在對前述資料集進行查詢之前或期間對分類或預分類資料集之需要。本發明之態樣使用例如前述元合併及/或元分組的索引來預組織前述資料，以使得當針對前述資料集執行查詢時，針對前述元合併及/或元分組的索引來執行前述查詢。

【英文】

Embodiments of the invention improve over conventional or routine technologies by generating a separate data structure or organization, other than those temporary search files that are typically used, that facilitates the processing of the datasets. In

one embodiment, instead of using sorted temp files or running sorting algorithms, a meta-join and/or meta-group-by indexes may be generated or created that provide a quick access to the records in the datasets. Moreover, embodiments of the invention eliminate the need to sort or pre-sort datasets before or during a query is conducted on the dataset. Aspects of the invention use, for example, the meta-join and/or meta-group-by indexes to pre-organize the data such that, when executing a query against the dataset, the query is executed against the meta-join and/or meta-group-by indexes.

【代表圖】

【本案指定代表圖】：圖 6。

【本代表圖之符號簡單說明】：

602 方塊

604 方塊

606 方塊

608 方塊

610 方塊

612 方塊

614 方塊

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

無。

圖式

100

102

#	LUID	PID	年齡	性別	收入
1	L1	P1	50 <u>106</u>	F <u>104</u>	60K
2	L1	P2	35	F	65K
3	L2	P3	70	M	120K
4	L2	P4	70	F	100K
5	L3	P5	50	F	200K
6	L3	P6	35	M	160K
7	L4	P7	50	F	80K
8	L5	P8	35	M	220K
9	L5	P9	35	F	100K
10	L5	P10	6	M	0

圖 1

#	LUID	Network	Series	ViewedFor	ShowDate
1	L1	Discovery	Life	45	01/03/2016
2	L2	ABC	Twisted	60	01/03/2016
3	L3	CNN	Early Start	30	01/03/2016
4	L4	Discovery	Life	45	01/03/2016
5	L5	Disney	Mickey Mouse	20	01/03/2016
6	L1	ABC	Twisted	45	01/03/2016
7	L2	Discovery	Life	60	01/03/2016
8	L3	Discovery	Life	20	01/03/2016
9	L4	Disney	Mickey Mouse	45	01/03/2016
10	L1	CNN	Wolf	45	01/03/2016
11	L2	Discovery	Life	50	01/04/2016
12	L3	Discovery	Life	50	01/04/2016
13	L1	Discovery	Life	60	01/04/2016
14	L2	CNN	Early Start	20	01/04/2016
15	L5	Discovery	Life	30	01/05/2016
16	L4	Disney	Mickey Mouse	45	01/05/2016
17	L3	Discovery	Life	45	01/05/2016
18	L2	CNN	Wolf	60	01/05/2016
19	L1	CNN	Early Start	60	01/06/2016
20	L1	CNN	Wolf	20	01/06/2016
21	L2	ABC	Twisted	30	01/06/2016
22	L3	ABC	Twisted	30	01/06/2016
23	L4	CNN	Wolf	45	01/06/2016
24	L5	CNN	Wolf	45	01/06/2016
25	L5	ABC	Twisted	30	01/06/2016
26	L4	Disney	Mickey Mouse	20	01/06/2016

圖2

300

302 304 306

LUID	消費者觀看	TV觀眾
L1	1,2 102	1,6,10,13,19,20 202
L2	3,4	2,7,11,14,18,21
L3	5,6	3,8,12,17,22
L4	7	4,9,16,23,26
L5	8,9,10	5,15,24,25

圖3

400

402

404

#	Gender=F	Gender=M	AGE=70	AGE=50	AGE=35	AGE=6
1	1	0	0	1	0	0
2	1	0	0	0	1	0
3	0	1	1	0	0	0
4	1	0	1	0	0	0
5	1	0	0	1	0	0
6	0	1	0	0	1	0
7	1	0	0	1	0	0
8	0	1	0	0	1	0
9	1	0	0	0	1	0
10	0	1	0	0	0	1

圖4

500

502

504

#	NW=ABC	NW=CNN	NW=DISCOVERY	NW=DISNEY	SER=LIFE
1	0	0	1	0	1
2	1	0	0	0	0
3	0	1	0	0	0
4	0	0	1	0	1
5	0	0	0	1	0
6	1	0	0	0	0
7	0	0	1	0	1
8	0	0	1	0	1
9	0	0	0	1	0
10	0	1	0	0	0
11	0	0	1	0	1
12	0	0	1	0	1
13	0	0	1	0	1
14	0	1	0	0	0
15	0	0	1	0	1
16	0	0	0	1	0
17	0	0	1	0	1
18	0	1	0	0	0
19	0	1	0	0	0
20	0	1	0	0	0
21	1	0	0	0	0
22	1	0	0	0	0
23	0	1	0	0	0
24	0	1	0	0	0
25	1	0	0	0	0
26	0	0	0	1	0

圖5

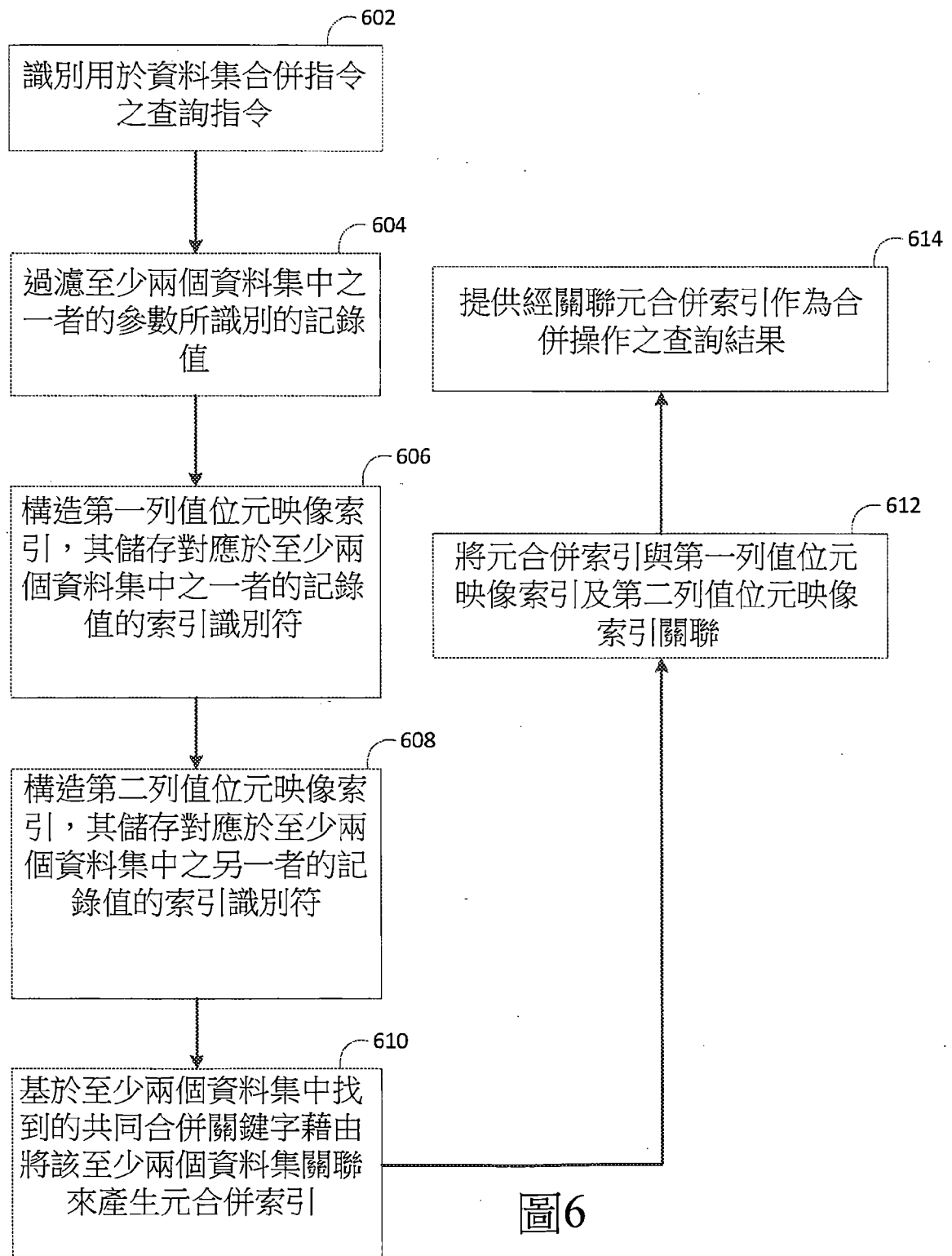


圖6

700

702

704

706

	消費者觀看位置	TV觀眾位置
L1	1	1
	1	13
	2	1
	2	13
L2	*	*
	*	*

圖7

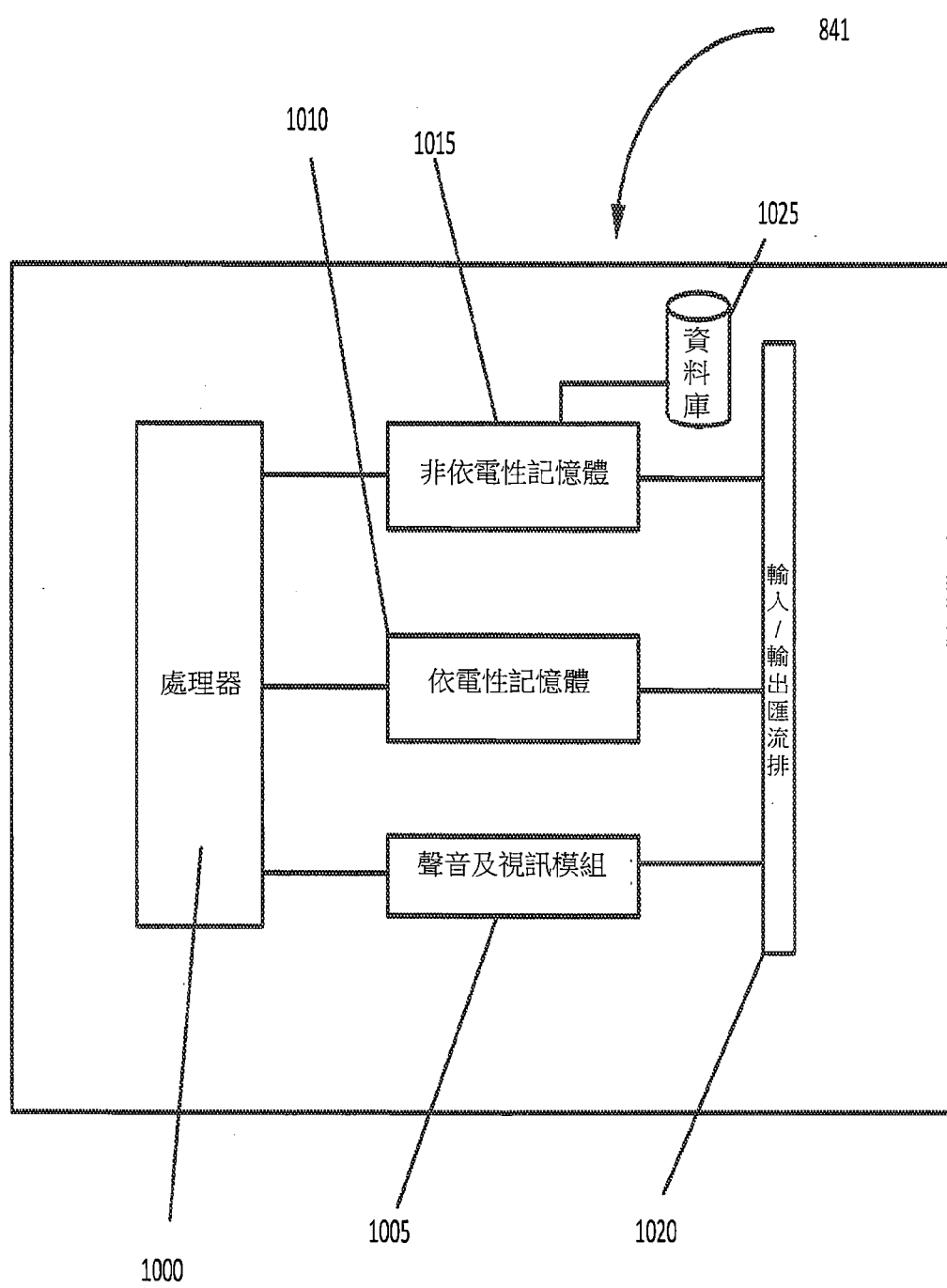


圖8

900

Network-Series	消費者觀看位置	TV觀眾位置
Discovery-Life	1	1
	1	13
	2	1
	2	13
	3	7
	3	11
	4	7
	4	11
	5	8
	5	12
	5	17
	6	8
	6	12
	6	17
	7	4
	8	15
	9	15
	10	15
ABC-Twisted	1	6
	2	6
	3	2
	*	*

圖9

1002

Network+Series
Discovery-Life
ABC-Twisted
CNN-Early Start
Disney-Mickey Mouse
CNN-Wolf

圖 10

1100

合併關鍵字	消費者觀看 合併關鍵字位置	TV觀眾 合併關鍵字位置	TV觀眾 分組位置
L1	1,2	1,6,10,13,19,20 <u>1102</u>	1, 13 <u>1104</u>
L2	3,4	2,7,11,14,18,21	7,11 <u>1106</u>
L3	5,6	3,8,12,17,22	8,12,17
L4	7	4,9,16,23,26	4
L5	8,9,10	5,15,24,25	15

圖 11

1200

合併關鍵字	消費者觀看 合併關鍵字位置	TV觀眾 合併關鍵字位置	TV觀眾 分組位置
L1	1,2	1,6,10,13,19,20	6 <u>1202</u>
L2	3,4	2,7,11,14,18,21	2,21 <u>1204</u>
L3	5,6	3,8,12,17,22	22
L4	7	4,9,16,23,26	
L5	8,9,10	5,15,24,25	25

圖12

1300

Network+Series	消費者觀看位置	TV觀眾位置
Discovery-Life	1	1
	1	13
	2	1
	2	13
	3	7
	3	11
	4	7
	4	11
	5	8
	5	12
	5	17
	6	8
	6	12
	6	17
	7	4
	8	15
	9	15
	10	15
ABC-Twisted	1	6
	2	6
	3	2
	3	21
	4	2
	4	21
	5	22
	6	22
	8	25
	9	25
	10	25

圖13

1400

Network+Series	消費者觀看位置	TV觀眾位置
Discovery-Life	1	1
	1	13
	2	1
	2	13
	4	7
	4	11
	5	8
	5	12
	5	17
	7	4
	9	15

圖14

1500

1502 1504 1506

Network+Series	消費者觀看位置	TV觀眾位置
Discovery-Life	1	1
	1	13
	2	1
	2	13
	4	7
	4	11
	5	12
	5	17
	7	4

圖15

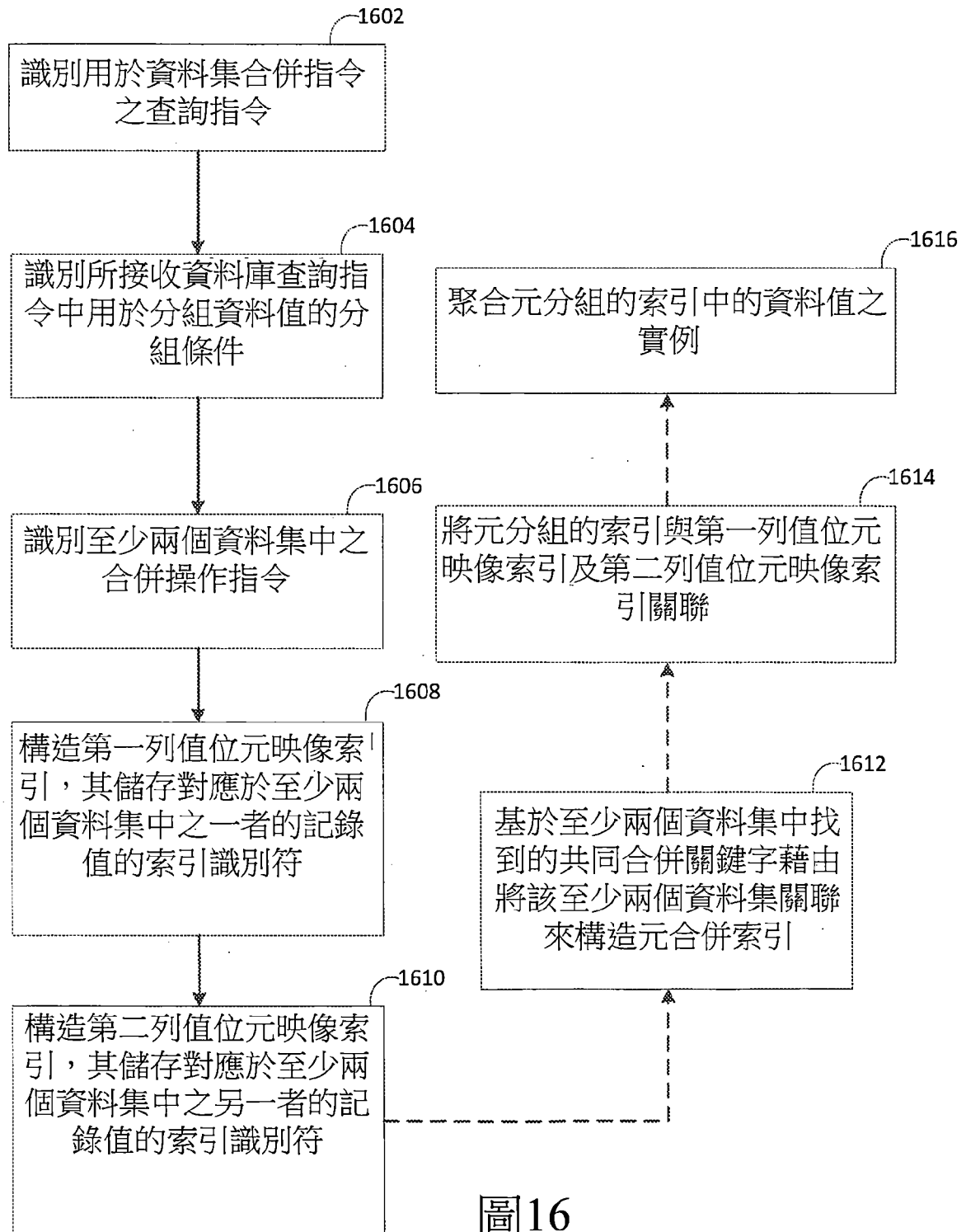


圖 16

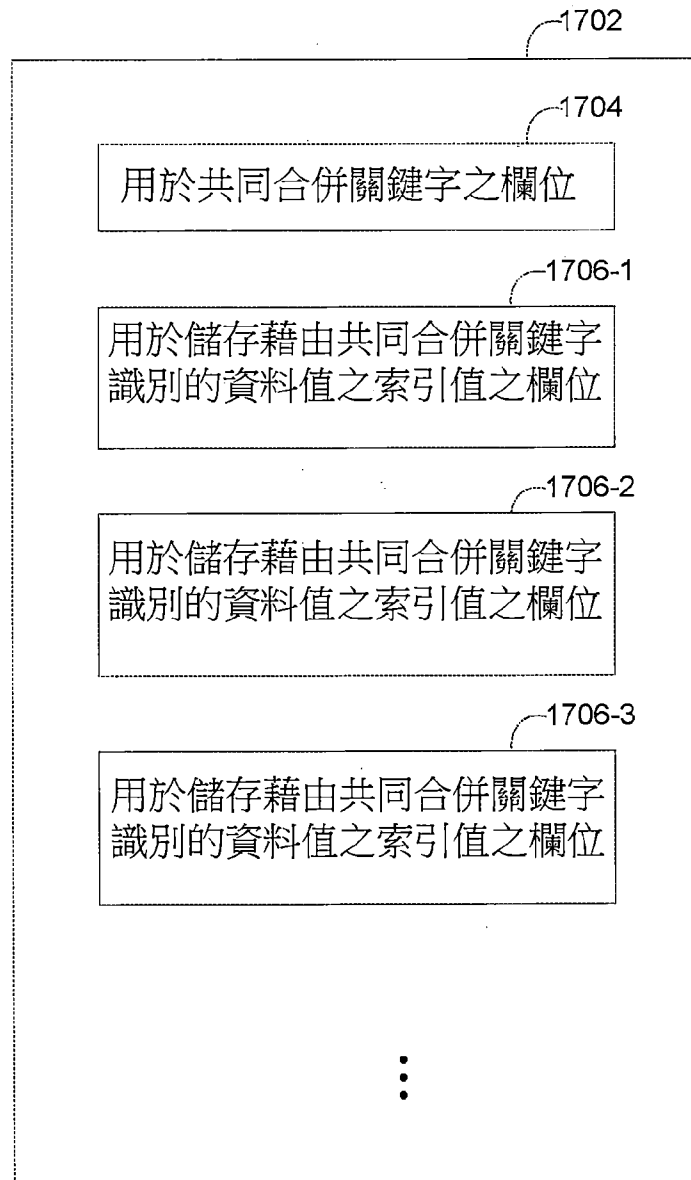


圖17

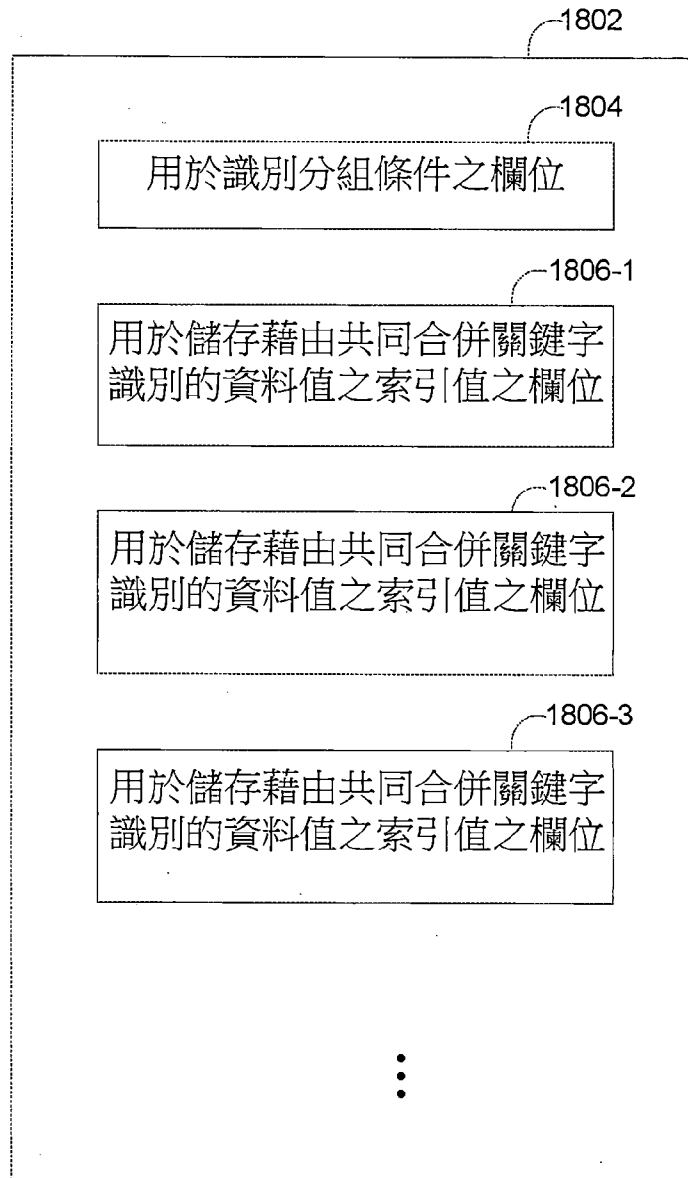


圖18

申請專利範圍

1. 一種電腦化方法，係用於產生複數個表示用於儲存在非暫時性電腦可讀媒體上之元合併索引的資料欄位，前述電腦化方法包含以下步驟：

藉由電腦可執行處理器識別用於複數個資料集中之至少兩個資料集的合併操作指令；

在不對前述至少兩個資料集執行前述合併操作指令的情況下，藉由前述電腦可執行處理器回應於前述經識別合併操作指令的要求並根據合併操作指令中之參數創建元合併索引，前述創建前述元合併索引包含：

藉由前述電腦可執行處理器過濾藉由在前述合併操作指令中之參數識別的前述至少兩個資料集中之一者中的記錄值；

藉由前述電腦可執行處理器構造第一列值位元映象索引，其儲存對應於藉由前述參數識別的前述至少兩個資料集之一者中的前述經過濾記錄值的索引識別符；

藉由前述電腦可執行處理器構造第二列值位元映象索引，其儲存對應於藉由前述參數識別的前述至少兩個資料集之另一者中的前述經過濾記錄值的索引識別符；

藉由前述電腦可執行處理器，藉由將前述至少兩個資料集基於在前述至少兩個資料集中找到的

共同合併-關鍵字關聯來產生前述元合併索引；

以及

藉由前述電腦可執行處理器對前述經創建元合併索引執行前述合併操作指令。

2. 如請求項 1 所記載之電腦化方法，其中前述共同合併-關鍵字包含局部唯一識別符(LUID)。
3. 如請求項 1 所記載之電腦化方法，其中前述元合併索引包含資料結構，其具有用於識別前述共同合併-關鍵字之資料欄位及複數個各自用於儲存索引值之資料欄位。
4. 如請求項 1 所記載之電腦化方法，其中創建前述元合併索引包含創建前述元合併索引而不分類前述至少兩個資料集。
5. 如請求項 1 所記載之電腦化方法，其中創建前述元合併索引包含創建前述元合併索引而不創建用於前述至少兩個資料集之臨時表格。
6. 如請求項 1 所記載之電腦化方法，其中創建前述元合併索引包含創建前述元合併索引而不進行對前述至少兩個資料集之檔案尋覓操作。
7. 一種電腦化系統，係用於產生複數個表示用於儲存在非暫時性電腦可讀媒體上之元合併索引的資料欄位，前述電腦化系統包含：
 - 用於儲存資料集之記憶體；
 - 可存取前述記憶體之電腦可執行處理器，其經組

配用於執行電腦可執行指令以供：

自使用者接收用於合併至少兩個資料集中之資料的資料庫查詢指令；

識別用於複數個資料集中之至少兩個資料集之合併操作指令；

在未執行對前述記憶體中之前述至少兩個資料集執行前述合併操作指令，回應於前述經識別合併操作指令的要求並根據合併操作指令中之參數創建元合併索引，前述創建前述元合併索引包含：

過濾藉由在前述合併操作指令中之參數識別的前述至少兩個資料集中之一者中的記錄值；

構造第一列值位元映象索引，其儲存對應於藉由前述參數識別的前述至少兩個資料集之一者中的前述經過濾記錄值的索引識別符；

構造第二列值位元映象索引，其儲存對應於藉由前述參數識別的前述至少兩個資料集之另一者中的前述經過濾記錄值的索引識別符；

藉由將前述至少兩個資料集基於在前述至少兩個資料集中找到的共同合併-關鍵字關聯來產生前述元合併索引；

以及

對前述經創建元合併索引執行前述合併操作指令。

8. 如請求項 7 所記載之電腦化系統，其中前述共同合併

-關鍵字包含局部唯一識別符(LUID)。

9. 如請求項 7 所記載之電腦化系統，其中前述元合併索引包含表格。
10. 如請求項 7 所記載之電腦化系統，其中前述處理器經組配以創建前述元合併索引包含前述處理器經組配以創建前述元合併索引而不首先分類前述至少兩個資料集。
11. 如請求項 7 所記載之電腦化系統，其中前述處理器經組配以創建前述元合併索引包含前述處理器經組配以創建前述元合併索引而不創建用於前述至少兩個資料集之臨時表格。
12. 如請求項 7 所記載之電腦化系統，其中前述處理器經組配以創建前述元合併索引包含前述處理器經組配以創建前述元合併索引而不對前述至少兩個資料集執行檔案尋覓操作。
13. 一種電腦化系統，係用於產生複數個表示用於儲存在非暫時性電腦可讀媒體上之元分組的索引的資料欄位，前述電腦化系統包含：
 - 用於儲存資料集之記憶體；
 - 可存取前述記憶體之電腦可執行處理器，其經組配用於執行電腦可執行指令以供：
 - 接收用於分組至少兩個資料集中之資料的資料庫查詢指令；
 - 識別前述所接收資料庫查詢指令中用於分組

前述至少兩個資料集中之資料值的分組條件；

在未執行對前述記憶體中之前述至少兩個資料集執行前述分組條件，回應於前述經識別分組條件的要求並根據分組條件中之參數創建元分組的索引，前述創建前述元分組的索引包含：

識別用於前述至少兩個資料集之合併操作指令；

過濾藉由在前述合併操作指令中之參數識別的前述至少兩個資料集中之一者中的記錄值；

構造第一列值位元映象索引，其儲存對應於藉由前述經識別參數識別的前述至少兩個資料集之一者中的前述記錄值的索引識別符；

構造第二列值位元映象索引，其儲存對應於藉由前述經識別參數識別的前述至少兩個資料集之另一者中的前述記錄值的索引識別符；

以及

藉由前述電腦可執行處理器對前述經創建元分組的索引執行前述分組條件。

14. 如請求項 13 所記載之電腦化系統，其中前述處理器經進一步組配以在構造前述第二列值位元映象索引之後執行以下電腦可執行指令：

藉由將前述至少兩個資料集基於在前述至少兩個資料集中找到的共同合併-關鍵字關聯來構造元合併索引；以及

將前述元合併索引與前述第一列值位元映像索引及前述第二列值位元映像索引關聯。

15. 如請求項 13 所記載之電腦化系統，其中前述處理器經進一步組配以聚合前述元分組的索引中之資料值之實例。
16. 如請求項 13 所記載之電腦化系統，其中前述共同合併-關鍵字包含局部唯一識別符(LUID)。
17. 如請求項 13 所記載之電腦化系統，其中前述元合併索引包含表格。
18. 如請求項 13 所記載之電腦化系統，其中前述處理器經組配以創建前述元合併索引包含前述處理器經組配以創建前述元合併索引而不首先分類前述至少兩個資料集。
19. 如請求項 13 所記載之電腦化系統，其中前述處理器經組配以創建前述元合併索引包含前述處理器經組配以創建前述元合併索引而不創建用於前述至少兩個資料集之臨時表格。
20. 如請求項 13 所記載之電腦化系統，其中前述處理器經組配以創建前述元合併索引包含前述處理器經組配以創建前述元合併索引而不對前述至少兩個資料集執行檔案尋覓操作。