



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 201421387 A

(43)公開日：中華民國 103 (2014) 年 06 月 01 日

(21)申請案號：102143382

(22)申請日：中華民國 102 (2013) 年 11 月 28 日

(51)Int. Cl.：

G06Q10/10 (2012.01)

G06F17/30 (2006.01)

G06F21/00 (2013.01)

(30)優先權：2012/11/30

日本

2012-263086

(71)申請人：U B I C 股份有限公司 (日本) UBIC, INC. (JP)

日本

(72)發明人：守本正宏 MORIMOTO, MASAHIRO (JP)；白井喜勝 SHIRAI, YOSHIKATSU

(JP)；武田秀樹 TAKEDA, HIDEKI (JP)；蓮子和巳 HASUKO, KAZUMI (JP)

(74)代理人：李保祿

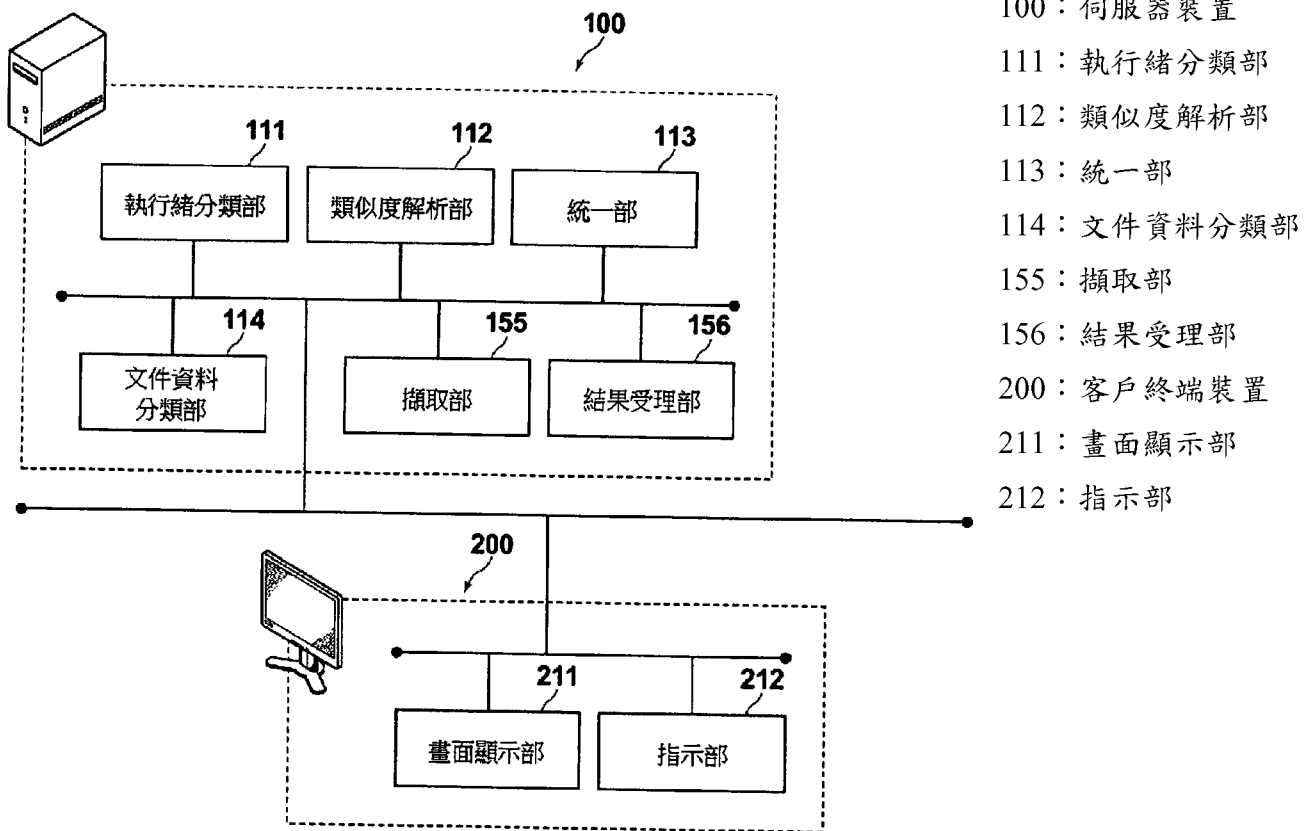
申請實體審查：無 申請專利範圍項數：9 項 圖式數：7 共 31 頁

(54)名稱

文件管理系統及文件管理方法以及文件管理程式

(57)摘要

本發明可減輕檢閱者檢閱的負荷。本發明係取得記錄於複數台電腦或伺服器的數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析之文件管理系統，其具備：執行緒分類部，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部，其係基於此類似度而統一執行緒彼此。



第 1 圖



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 201421387 A

(43)公開日：中華民國 103 (2014) 年 06 月 01 日

(21)申請案號：102143382

(22)申請日：中華民國 102 (2013) 年 11 月 28 日

(51)Int. Cl.：

G06Q10/10 (2012.01)

G06F17/30 (2006.01)

G06F21/00 (2013.01)

(30)優先權：2012/11/30

日本

2012-263086

(71)申請人：U B I C 股份有限公司 (日本) UBIC, INC. (JP)

日本

(72)發明人：守本正宏 MORIMOTO, MASAHIRO (JP)；白井喜勝 SHIRAI, YOSHIKATSU

(JP)；武田秀樹 TAKEDA, HIDEKI (JP)；蓮子和巳 HASUKO, KAZUMI (JP)

(74)代理人：李保祿

申請實體審查：無 申請專利範圍項數：9 項 圖式數：7 共 31 頁

(54)名稱

文件管理系統及文件管理方法以及文件管理程式

(57)摘要

本發明可減輕檢閱者檢閱的負荷。本發明係取得記錄於複數台電腦或伺服器的數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析之文件管理系統，其具備：執行緒分類部，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部，其係基於此類似度而統一執行緒彼此。

發明摘要

※ 申請案號：

※ 申請日：102143382

102. 11. 28

※IPC 分類：

G06Q 10/10 (2012.01)
G06F 17/30 (2006.01)
G06F 21/00 (2013.01)

【發明名稱】文件管理系統及文件管理方法以及文件管理程式

【中文】

本發明可減輕檢閱者檢閱的負荷。

本發明係取得記錄於複數台電腦或伺服器的數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析之文件管理系統，其具備：執行緒分類部，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部，其係基於此類似度而統一執行緒彼此。

【英文】

【代表圖】

【本案指定代表圖】：第（ 1 ）圖。

【本代表圖之符號簡單說明】：

- 100 伺服器裝置
- 111 執行緒分類部
- 112 類似度解析部
- 113 統一部
- 114 文件資料分類部
- 155 擷取部
- 156 結果受理部
- 200 客戶終端裝置
- 211 畫面顯示部
- 212 指示部

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

發明專利說明書

(本說明書格式、順序，請勿任意更動)

【發明名稱】 文件管理系統及文件管理方法以及文件管理程式

【技術領域】

【0001】 本發明係關於一種文件管理系統及文件管理方法以及文件管理程式，特別是關於一種用於收集與訴訟相關之文件資訊的文件管理系統及文件管理方法以及文件管理程式。

【先前技術】

【0002】 以往提出了產生不正當存取或機密資訊洩漏等關於電腦的犯罪或法律上的紛爭之際，追究原因或搜查所需的機器或收集分析資料、電子式的記錄，查明其法律上的證據性之手段或技術。

【0003】 此外，在美國民事訴訟中被要求 eDiscovery(電子證據公開)等，該訴訟的原告及被告都要承擔提出全部相關的數位資訊作為證據的責任。因此，必須提出記錄於電腦或伺服器的數位資訊作為證據。

【0004】 另一方面，伴隨 IT 的迅速發展與普及，在現代的商務世界中都以電腦製作大部分的資訊，所以即使是同一企業內也是許多的數位資訊充斥著。

【0005】 因此，在進行用於向法庭提出證據資料的準備作業之過程中，容易產生連與該訴訟未必相關的機密性數位資訊都包含作為證據資料的失誤。此外，提出與該訴訟無關的機密性文件資訊成為問題。

【0006】 近幾年，專利文獻 1 至專利文獻 3 上提出了關於文件管理系

統的文件資訊之技術。專利文獻 1 上揭示了一種文件管理系統：從文件提出命令的對象者資訊中所含的至少一人以上的對象者中指定特定者，基於關於所指定之特定者的存取履歷資訊，僅擷取特定者存取的數位文件資訊，設定顯示所擷取的數位文件資訊之各文件檔案是否是與訴訟相關的附帶資訊，基於附帶資訊，輸出與訴訟相關的文件檔案。

【0007】 此外，專利文獻 2 上揭示了一種文件管理系統：顯示所記錄的數位資訊，對複數個各文件檔案都設定顯示與對象者資訊中所含的對象者之中哪個對象者相關的對象者特定資訊，並設定成在記憶部內記錄該所設定的對象者特定資訊，指定至少一人以上的對象者，檢索設定有對應於所指定之對象者的對象者特定資訊之文件檔案，經由顯示部而設定顯示所檢索的文件檔案是否是與訴訟相關的附帶資訊，基於附帶資訊，輸出與訴訟相關的文件檔案。

【0008】 再者，專利文獻 3 上揭示了一種文件管理系統：受理數位文件資訊中所含的至少一個以上的文件檔案的指定，受理將所指定的文件檔案翻譯成哪種語言的指定，將受理指定的文件檔案翻譯成受理指定的語言，從記錄於記憶部的數位文件資訊中擷取顯示和所指定的文件檔案同一內容的共通文件檔案，生成顯示藉由引用所翻譯的文件檔案之翻譯內容而翻譯所擷取的共通文件檔案之翻譯相關資訊，基於翻譯相關資訊，輸出與訴訟相關的文件檔案。

【0009】

【先前技術文獻】

【專利文獻】

【專利文獻 1】特開 2011－209930 號公報

【專利文獻 2】特開 2011－209931 號公報

【專利文獻 3】特開 2012－32859 號公報

【發明內容】

【發明欲解決之課題】

【0010】 然而，在例如專利文獻 1 至專利文獻 3 之類的文件管理系統方面，要收集利用複數台電腦及伺服器之對象者的龐大文件資訊。

【0011】 進行將此種經數位化的龐大文件資訊作為訴訟的證據資料是否妥當的分類之作業，需要被稱為檢閱者的利用者利用目視確認，一個一個逐漸分類該文件資訊，有花費很大的勞力與時間的傾向，需要進一步的改善。

【0012】 因此，本發明有鑑於上述情況，其目的在於提供一種可更加減輕檢閱者檢閱的負荷之文件管理系統及文件管理方法以及文件管理程式。

【解決課題之手段】

【0013】 本發明之文件管理系統，係取得記錄於複數台電腦或伺服器的數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析，其具備：執行緒分類部，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部，其係基於此類似度而統一執

行緒彼此。

【0014】 「附帶資訊」係指附隨於文件資料的資訊。例如，附帶資訊也可以是為將文件資料構造化而附加標籤的元資訊等。

【0015】 「執行緒」係指關於某特定資訊的文件資料的集合。例如，執行緒分類部可將就共通的内容所記載的電子郵件及對於該電子郵件的回覆郵件作為一個執行緒而將文件資料進行分類。

【0016】 此外，上述文件管理系統也可以進一步具備：文件資料分類部，其係將數位資訊中所含的文件資料分類為單獨文件資料、主文件資料、和主文件資料有從屬關係的從屬文件資料、及其他的文件資料；上述執行緒分類部至少將單獨文件資料及主文件資料進行分類。

【0017】 此外，上述執行緒分類部也可以進一步將從屬文件資料之各個文件資料分類為和對應的主文件資料相同的執行緒。

【0018】 此外，執行緒分類部也可以處理文件資料的標題、作成日期和時間、更新日期和時間、作者名之中至少任一個作為附帶資訊。

【0019】 此外，其他的文件資料也可以包含沒有附帶資訊的非構造化文件資料，執行緒分類部也可以將非構造化文件資料進行分類之際，參考該非構造化文件資料的内容。

【0020】 此外，上述統一部也可以將從屬文件資料之各個文件資料統一為和對應的主文件資料相同的執行緒。

【0021】 此外，上述文件管理系統也可以進一步具備：擷取部，其係從所統一的執行緒擷取預定數量的文件資料；畫面顯示部，其係在畫面上顯示此等所擷取的文件資料；結果受理部，其係受理利用者對於所顯示的

文件資料進行的相關性判斷結果；選定部，其係基於判斷結果，將所擷取的文件資料分類為各判斷結果，在此等所分類的文件資料中解析、選定共通出現的要素；要素記錄部，其係記錄選定的要素；檢索部，其係從文件資料檢索記錄於要素記錄部的要素；及分數計算部，其係使用檢索部的檢索結果與選定部的解析結果，計算表示判斷結果與文件資料相關性的分數。

【0022】 本發明之文件管理方法，係取得記錄於複數台電腦或伺服器之數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析，電腦執行以下步驟：確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及基於此類似度而統一執行緒彼此。

【0023】 本發明之文件管理程式，係取得記錄於複數台電腦或伺服器之數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析，使電腦實現以下功能：確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及基於此類似度而統一執行緒彼此。

【發明之效果】

【0024】 本發明之文件管理系統，係取得記錄於複數台電腦或伺服器之數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析，藉由具備：執行緒分類部，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部，其係將所

分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部，其係基於此類似度而統一執行緒彼此；當利用者進行相關性判斷(檢閱)之際，可用同質的文件資料單位進行檢閱，可實現檢閱負荷的減輕及效率化。

【0025】 此外，上述文件管理系統於下述構造之際：進一步具備：文件資料分類部，其係將數位資訊中所含的文件資料分類為單獨文件資料、主文件資料、和主文件資料有從屬關係的從屬文件資料、及其他的文件資料，上述執行緒分類部至少將單獨文件資料及主文件資料進行分類；可至少將單獨文件資料及主文件資料分類為執行緒。

【0026】 此外，上述統一部於下述構造之際：將從屬文件資料之各個文件資料統一為和對應的主文件資料相同的執行緒；可使有從屬關係的文件資料統一為相同的執行緒。

【0027】 此外，上述執行緒分類部於下述構造之際：將從屬文件資料之各個文件資料分類為和對應的主文件資料相同的執行緒；可將有從屬關係的文件資料分類為相同的執行緒。

【0028】 此外，執行緒分類部於下述構造之際：處理文件資料的標題、作成日期和時間、更新日期和時間、作者名之中至少任一個作為附帶資訊；可基於文件資料內容以外的資訊而將文件資料分類為執行緒。

【0029】 此外，於下述構造之際：其他的文件資料包含沒有附帶資訊的非構造化文件資料，執行緒分類部將非構造化文件資料進行分類之際，參考該非構造化文件資料的內容；即使是沒有附帶資訊的文件資料，也可以利用代替手段分類為執行緒。

【0030】 此外，上述文件管理系統於下述構造之際：進一步具備：擷取部，其係從所統一的執行緒擷取預定數量的文件資料；畫面顯示部，其係在畫面上顯示此等所擷取的文件資料；結果受理部，其係受理利用者對於所顯示的文件資料進行的相關性判斷結果；選定部，其係基於判斷結果，將所擷取的文件資料分類為各判斷結果，在此等所分類的文件資料中解析、選定共通出現的要素；要素記錄部，其係記錄選定的要素；檢索部，其係從文件資料檢索記錄於要素記錄部的要素；及分數計算部，其係使用檢索部的檢索結果與選定部的解析結果，計算表示判斷結果與文件資料相關性的分數；對於同質的文件資料，依據利用者的檢閱結果，系統可自動進行檢閱，可使檢閱的精度提高。

【圖式簡單說明】

【0031】

第 1 圖為本發明第一實施形態之文件管理系統的方塊圖；

第 2 圖為顯示本發明第一實施形態之處理流程的流程圖；

第 3 圖為顯示在本發明第一實施形態中利用處理將文件資料進行分類之情況的概略圖；

第 4 圖為顯示本發明第一實施形態之電子郵件親子關係的概略圖；

第 5 圖為本發明第二實施形態之文件管理系統的方塊圖；

第 6 圖為顯示本發明第二實施形態之選定部解析結果的圖表；以及

第 7 圖為顯示本發明第二實施形態之處理流程的流程圖。

【實施方式】**【0032】**〔第 1 實施形態〕

以下，使用第 1 圖至第 4 圖說明本發明之第一實施形態。

【0033】 第 1 圖顯示關於本實施形態之文件管理系統的方塊圖。

文件管理系統具備電腦或伺服器，基於各種輸入，CPU 執行記錄於 ROM 的程式，而動作作為各種功能部。該程式也可以是記憶於 CD-ROM 等記憶媒體，或者經由網際網路等網路而分發，安裝於電腦上者。

【0034】 關於本發明第一實施形態之文件管理系統係由客戶終端裝置 200 及伺服器裝置 100 所構成。客戶終端裝置 200 具備：畫面顯示部 211，其係顯示成為利用者就和訴訟的相關性進行判斷之對象的由數位資訊擷取之複數個文件資料；及指示部 212，其係受理來自利用者的輸入。

【0035】 伺服器裝置 100 具備：執行緒分類部 111，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部 112，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部 113，其係基於此類似度而統一執行緒彼此。進一步具備：文件資料分類部 114，其係將數位資訊中所含的文件資料分類為單獨文件資料、主文件資料、和該主文件資料有從屬關係的從屬文件資料、及其他的文件資料。此時，執行緒分類部 111 至少將單獨文件資料及主文件資料進行分類。伺服器裝置進一步具備：擷取部 155，其係從上述所統一的執行緒擷取預定數量的文件資料；及結果受理部 156，其係受理利用者對於畫面顯示部 211 所顯示的文件資料進行的相關性判斷結果。

【0036】 在 eDiscovery(電子證據公開)方面，需要進行將經數位化的龐大文件資訊作為訴訟的證據資料是否妥當的分類之作業(稱為檢閱)。檢閱(相關性判斷)係指系統或人對於文件資料判斷有無對訴訟的提出。例如，在檢閱方面，將成為檢閱對象的文件資料即文件群基於訴訟相關的程度或與

訴訟相關的做法而進行分類成複數種類。

【0037】 在關於本實施形態之文件管理系統方面，係基於預定的要件作成文件資料的集合作為進行此檢閱前的階段。然後，被稱為檢閱者的利用者為了對於所匯集的文件資料從文件資料中擷取在訴訟中必須提出的文件資料，而進行和訴訟的相關性判斷。藉此，檢閱者可對於同質的文件資料進行檢閱，可謀求檢閱的效率化及檢閱者負擔的減輕。

【0038】 文件資料係指包含一個以上的單字的資訊。文件資料為具有包含一個以上的單字的資訊作為本文的內容，具有附帶資訊作為本文以外的資訊者較好。所謂附帶資訊係指元資訊。例如，附帶資訊也可以意味著文件資料的作成日期和時間或作成者、資料形式、標題等與文件資料相關的資訊。

【0039】 作為文件資料的一例，可舉電子郵件、提交資料、試算表資料、協商資料、契約書、組織圖、事業計劃書等。此外，也可以操作掃描資料作為文件資料。此情況，也可以在文件管理系統內具備 OCR(Optical Character Reader；光學字元閱讀器)裝置，以便可將掃描資料轉換為文字資料。利用 OCR 裝置變更為文字資料，可從掃描資料中解析或檢索後述的要素。文件資料也可以是藉由後述的文件資料分類部 114 的處理而分類為單獨文件資料、主文件資料、從屬文件資料、其他的文件資料者。

【0040】 茲就第 1 圖所記載的各構造的功能進行說明。

【0041】 伺服器裝置 100 與客戶終端裝置 200 係經由通信網路而連接。通信網路係指有線或無線的通信線路。例如，為電話線路、網際網路線路等。

【0042】 畫面顯示部 211 係指在客戶終端裝置 200 上顯示成為檢閱對象的文件資料之顯示器或液晶監視器等。指示部 212 係指配備於客戶終端裝置 200 的滑鼠或鍵盤，受理來自利用者的操作或輸入。

【0043】 文件資料分類部 114 將文件資料分類為單獨文件資料、主文

件資料、和主文件資料有從屬關係的從屬文件資料、及其他的文件資料。所謂從屬文件資料，係指如同電子郵件的附加檔案，具有和主文件資料(此情況為電子郵件)的從屬關係者。單獨文件資料係指沒有從屬文件資料的文件資料。其他的文件資料係指都不符合單獨文件資料、主文件資料及從屬文件資料之任一者的文件資料。例如，非構造化文件資料等相當於此文件資料。非構造化文件資料意味著沒有附帶資訊的文件資料。具體而言，件名為空白的電子郵件相當於此文件資料。

【0044】 執行緒分類部 111 將文件資料基於該等附帶資訊而分類為執行緒。例如，將電子郵件進行分類時，執行緒分類部 111 基於電子郵件的件名，將同一件名者、或在同一件名上附加有表示回覆的「Re：」或表示轉寄的「FW：」之件名的電子郵件分類為同一執行緒。除了件名之外，執行緒分類部 111 也可以處理文件資料的標題、作成日期和時間、更新日期和時間、作者名之中至少任一個作為附帶資訊。此外，於將例如件名為空白的電子郵件之類的沒有附帶資訊的非構造化文件資料進行分類之際，執行緒分類部 111 也可以參考此等非構造化文件資料的內容。即，在將非構造化文件資料的電子郵件進行分類時，係基於此電子郵件的本文而進行分類。

【0045】 類似度解析部 112 解析執行緒間的類似度。特別是類似度解析部 112 基於要素而解析執行緒間的類似度較好，該要素係包含於各執行緒包含於其執行緒的電子郵件的附帶資訊。要素係指文件資料中所含的單字、記號、圖面等文件資料的構成要素。例如，要素也可以是指詞素之類的某種語言方面，分割到若再分解則不構成意義的程度而所擷取的音素匯集各個音素。具體而言，「將文件進行分類」此一文章的要素也可以是「文件」「分類」「進行」。

【0046】 統一部 113 基於執行緒中所含的文件資料的附帶資訊的類似度而統一執行緒彼此。此外，統一部 113 將從屬文件資料之各個文件資

料統一為和主文件資料相同的執行緒。將統一部 113 統一的執行緒作為群集。

【0047】 在本實施形態中，擷取部 155 從由統一部 113 統一的執行緒(群集)中所含的文件資料擷取預定數量的文件資料作為利用者的檢閱對象。此擷取也可以隨機或基於文件資料的更新日期和時間等屬性而進行。

【0048】 結果受理部 156 受理利用者對於各文件資料進行的檢閱的結果。利用者在檢閱中按照和訴訟的相關程度，對文件資料進行識別符號的給予。例如，給予和訴訟的相關性高且必須提出作為證據的文件資料「Responsive」符號，給予不要提出的文件資料「Non-Responsive」符號。再者，將「HOT」符號和「Responsive」符號一併給予已給予「Responsive」符號的文件資料之中，特別是和訴訟的相關性高且格外重要的證據的文件資料。此外，給予已給予「Responsive」符號的文件資料之中，和製品 A 相關性高的內容的文件資料「製品 A」符號，給予和製品 B 相關性高的內容的文件資料「製品 B」符號。此時，結果受理部 156 可受理各識別符號作為相關性判斷的結果。

【0049】 其次，使用第 2 圖及第 3 圖，就文件管理系統之處理流程進行說明。第 2 圖為顯示本實施形態的文件管理系統之處理流程的流程圖。此外，第 3 圖為概略顯示在進行第 2 圖所示的處理之際，將文件資料逐漸群集化之情況的圖。

【0050】 首先，在第 2 圖中，文件資料分類部 114 將文件資料進行分類(步驟 111)。具體而言，如第 3 圖所示，文件資料分類部 114 於收集到的全部文件資料 501(ALL Documents)之中，將電子郵件分類為 Mails 組 511，將附加在電子郵件上的檔案分類為 Attachments 組 512，將並不是電子郵件或其附加檔案之任一者的文件資料分類為 Other Documents 組 513。

【0051】 其次，執行緒分類部 111 將 Mails 組 511 的文件以附帶資訊為電子郵件的件名，基於此而分類為執行緒群 521(Threaded Clusters)或非執

行緒群 522(Non-Threaded Clusters)。例如，如第 4 圖所示，關於件名為「製品 A 的開發狀況」的電子郵件，設定為主郵件有 1 件、回覆有 2 件、轉寄有 1 件，而關於「製品 A 的出貨判定」之件名的電子郵件，設定為主郵件有 1 件、回覆有 3 件。再者，以下述情況為例來進行說明：作為僅主郵件的郵件，件名為「新商品的開發」、「製品 A 的企畫」、「關於 B 先生」的郵件及件名為空白的郵件有 1 件。此時，件名為「製品 A 的開發狀況」、「Re：製品 A 的開發狀況」、「Re：Re：製品 A 的開發狀況」及「FW：Re：Re：製品 A 的開發狀況」的電子郵件係作為具有同一附帶資訊而被分類為同一執行緒(第 2 圖、步驟 112)。另一方面，只有主郵件的電子郵件不被進行執行緒化，而被歸納為第 3 圖的非執行緒群 522。

【0052】 其次，統一部 113 基於執行緒間的類似度而統一執行緒群 521 的執行緒與非執行緒群 522 中所含的文件資料，作為執行緒群 531(Threaded-based Clusters)(步驟 113)。此時，統一部 113 合成屬於執行緒的文件資料的附帶資訊，生成代表執行緒的附帶資訊。具體而言，擷取附帶資訊中所含的要素，基於要素間的類似度的分布而生成合成向量。然後，將被分類為非執行緒群 522 的文件資料之中具有和合成的附帶資訊類似度高的附帶資訊之文件資料統一為此合成的附帶資訊之執行緒群 521 中的執行緒。

【0053】 再者，統一部 113 將被分類為 Attachments 組的從屬文件資料統一為執行緒，進行群集化為群集 541(Clusters)的處理(步驟 114)。最好統一為和附加有從屬文件資料的電子郵件相同的執行緒。

【0054】 此外，統一部 113 統一被分類為 Other Documents 組的文件資料(步驟 115)。從此等文件資料的檔案名(除了副檔名之外)作成附帶資訊向量，計算和在步驟 114 作成的合成向量的類似度。若最大類似度為臨界值以上，則統一為最類似的群集。關於以此處理未被統一的文件資料，則從整個該其他的文件資料構成向量，計算和剛才的合成向量的類似度。然

後，統一為最大類似度為臨界值以上的群集。以此處理也未被統一的其他文件資料則全部匯集作為其他的群集。

【0055】 擷取部 155 從如此作成的各群集擷取成為檢閱對象的文件資料(步驟 116)。擷取部 155 雖然利用隨機抽樣擷取文件資料，但也可以按作成日期和時間順序或作成者名順序擷取。

【0056】 然後，利用者對於所擷取的文件資料實施檢閱(步驟 117)，結果受理部 156 受理檢閱結果(步驟 118)。藉此，可對於同質的文件資料進行檢閱。

【0057】 〔第二實施形態〕

以下，使用第 5 圖至第 7 圖說明本發明之第二實施形態。

【0058】 第 5 圖顯示關於本實施形態之文件管理系統的方塊圖。

【0059】 關於本發明第二實施形態之文件管理系統係由客戶終端裝置 200 及伺服器裝置 100 所構成。客戶終端裝置 200 具備：畫面顯示部 211，其係顯示成為利用者就和訴訟的相關性進行判斷之對象的由數位資訊擷取之複數個文件資料；及指示部 212，其係受理來自利用者的輸入。

【0060】 此外，伺服器裝置 100 具備：執行緒分類部 111，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部 112，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部 113，其係基於此類似度而統一執行緒彼此。

【0061】 伺服器裝置 100 進一步具備：擷取部 155，其係從上述所統一的執行緒擷取預定數量的文件資料；結果受理部 156，其係受理利用者對於畫面顯示部 211 所顯示的文件資料進行的相關性判斷結果；選定部 157，其係基於此等判斷結果，將所擷取的文件資料分類為各判斷結果，在此等所分類的文件資料中解析、選定共通出現的要素；要素記錄部 158，其係記錄選定的要素；檢索部 159，其係從文件資料檢索記錄於要素記錄部 158 的

要素；及分數計算部 160，其係使用檢索部 159 的檢索結果與選定部 157 的解析結果，計算表示判斷結果與文件資料相關性的分數。

【0062】 選定部 157 選定要素。例如，選定部 157 也可以解析、選定進行過同一判斷結果的文件資料共通出現的要素。判斷結果係指顯示利用者對於文件群進行的和訴訟的相關性判斷的結果。判斷結果也可以是指表示利用者給予文件的和訴訟的相關性程度的分類符號。此外，選定部 157 也可以於選定要素之際，參考要素的評估值。

【0063】 評估值係指表示要素特徵的值。評估值也可以表示要素具有的傳達資訊量。傳達資訊量係指在概率論或資訊論方面，表示兩個概率變數相互依存尺度的量。具體而言，傳達資訊量也可以是表示對於包含其要素的文件的相關性判斷結果與和其要素的依存關係的尺度。

【0064】 第 6 圖為選定部 157 解析被判斷為有相關性的文件資料共通頻繁出現之要素的結果的圖表。在第 6 圖中，縱軸 R_hot 係表示包含由利用者判斷為有相關性的所有文件資料之中被選定作為與被判斷為有相關性的文件資料有關係的要素之要素且被判斷為有相關性的文件資料之比例。橫軸 R_all 係表示利用者實施過檢閱的所有文件資料之中包含由後述檢索部 159 檢索過的要素的文件資料之比例。在本實施形態中，在選定部 157 方面，係選定標繪於比直線 $R_hot = R_all$ 上部的要素作為被判斷為有相關性的文件資料共通的要素。

【0065】 要素記錄部 158 係指記錄要素者。也可以是資料庫。

【0066】 檢索部 159 係指從文件資料檢索要素者。

【0067】 分數計算部 160 係指計算文件資料的分數者。也可以是基於文件資料中所含的要素的評估值而計算分數者。評估值也可以是指基於文件資料中的要素出現頻率或傳達資訊量而計算，要素在某文件資料中發揮的資訊量。

【0068】 分數係指在某文件資料中表示和訴訟的相關程度者。分數係

基於文件資料中所含的要素而被算出。例如，也可以是越是含有必須在訴訟時提出較高的要素的文件資料，越是具有較高的分數。文件資料也可以是基於一定的要件而給予分數的初始值者。例如，也可以是利用文件資料中出現的要素與要素具有的評估值而計算初始分數者。

【0069】 分數計算部 160 可利用文件群中出現的要素與要素具有的評估值，從下式計算分數：

【0070】 【數 1】

$$Scr = \sum_{i=0}^N i * (m_i * wgt_i^2) / \sum_{i=0}^N i * wgt_i^2 \quad \cdot \cdot \cdot \quad (1)$$

m_i : 第*i*個要素的出現頻率
 wgt_i : 第*i*個要素的評估值

各要素具有的評估值係以該要素具有的傳達資訊量為基礎而決定。該評估值可利用下式而進行學習：

【0071】 【數 2】

$$wgt_{i,L} = \sqrt{wgt_{i,L-1}^2 + \gamma_L * wgt_{i,L-1}^2 - \vartheta} = \sqrt{wgt_{i,L-1}^2 + \sum_{l=1}^L (\gamma_l * wgt_{i,l-1}^2 - \vartheta)} \quad \cdot \cdot \cdot \quad (2)$$

$wgt_{i,0}$: 學習前的第*i*個選定要素的評估值(初始值)
 $wgt_{i,L}$: 第*L*次學習後的第*i*個選定要素的評估值
 γ_L : 第*L*次學習的學習參數
 ϑ : 學習效果的臨界值

【0072】 茲使用第 7 圖，就本實施形態的預測資訊生成處理的流程進行說明。首先，擷取部 155 從由統一部 113 統一的群集擷取預定數量的文件資料(步驟 211)。畫面顯示部 211 使所擷取的文件資料顯示於客戶終端裝置 200(步驟 212)。結果受理部 156 受理利用者相關性判斷的結果(步驟 213)，選定部 157 從利用者相關性判斷的結果解析文件資料，選定要素(步驟 214)。所選定的要素為要素記錄部 158 所記錄(步驟 215)。其次，檢索部 159 檢索由各文件資料記錄的要素，分數計算部 160 使用式(1)計算各文件資料的分數(步驟 216)。

【0073】 關於其他的構造、功能，和第一實施形態同樣。

【0074】 文件管理系統係取得記錄於複數台電腦或伺服器的數位資訊，將該所取得的數位資訊就和訴訟的相關性進行分析，藉由具備：執行緒分類部 111，其係確認數位資訊中所含的文件資料各個的附帶資訊，基於此等附帶資訊而將文件資料分類為執行緒；類似度解析部 112，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各執行緒，基於此等所擷取的要素而解析執行緒間的類似度；及統一部 113，其係基於此類似度而統一執行緒彼此；當利用者進行相關性判斷(檢閱)之際，可用同質的文件資料單位進行檢閱，可實現檢閱負荷的減輕及效率化。

【0075】 此外，上述文件管理系統於下述構造之際：進一步具備：文件資料分類部 114，其係將數位資訊中所含的文件資料分類為單獨文件資料、主文件資料、和主文件資料有從屬關係的從屬文件資料、及其他的文件資料，上述執行緒分類部 111 至少將單獨文件資料及主文件資料進行分類；可至少將單獨文件資料及主文件資料分類為執行緒。

【0076】 此外，上述統一部 113 於下述構造之際：將從屬文件資料之各個文件資料統一為和對應的主文件資料相同的執行緒；可使有從屬關係的文件資料統一為相同的執行緒。

【0077】 此外，上述執行緒分類部 111 於下述構造之際：將從屬文件資料之各個文件資料分類為和對應的主文件資料相同的執行緒；可將有從屬關係的文件資料分類為相同的執行緒。

【0078】 此外，執行緒分類部 111 於下述構造之際：處理文件資料的標題、作成日期和時間、更新日期和時間、作者名之中至少任一個作為附帶資訊；可基於文件資料內容以外的資訊而將文件資料分類為執行緒。

【0079】 此外，於下述構造之際：其他的文件資料包含沒有附帶資訊的非構造化文件資料，執行緒分類部 111 將非構造化文件資料進行分類之際，參考該非構造化文件資料的內容；即使是沒有附帶資訊的文件資料，也可以利用代替手段分類為執行緒。

【0080】 此外，上述文件管理系統於下述構造之際：進一步具備：擷取部 155，其係從所統一的執行緒擷取預定數量的文件資料；畫面顯示部 211，其係在畫面上顯示此等所擷取的文件資料；結果受理部 156，其係受理利用者對於所顯示的文件資料進行的相關性判斷結果；選定部 157，其係基於判斷結果，將所擷取的文件資料分類為各判斷結果，在此等所分類的文件資料中解析、選定共通出現的要素；要素記錄部 158，其係記錄選定的要素；檢索部 159，其係從文件資料檢索記錄於要素記錄部 158 的要素；及分數計算部 160，其係使用檢索部 159 的檢索結果與選定部 157 的解析結果，計算表示判斷結果與文件資料相關性的分數；對於同質的文件資料，依據利用者的檢閱結果，系統可自動進行檢閱，可使檢閱的精度提高。

【0081】 [其他實施形態]

以上，雖然就本發明之實施形態進行了敘述，但本發明並不受已述實施形態限定，可基於本發明之技術性思想進行各種變形及變更。

【0082】 例如，在已述實施形態中，執行緒分類部 111 係將文件資料之中的主文件資料與單獨文件資料分類為執行緒。然而，執行緒分類部 111 也可以進一步將從屬文件資料之各個文件資料分類為和對應的主文件資料相同的執行緒。

【符號說明】

【0083】

- 100 伺服器裝置
- 111 執行緒分類部
- 112 類似度解析部
- 113 統一部
- 114 文件資料分類部
- 155 擷取部
- 156 結果受理部

- 157 選定部
- 158 要素記錄部
- 159 檢索部
- 160 分數計算部
- 200 客戶終端裝置
- 211 畫面顯示部
- 212 指示部

【生物材料寄存】

國內寄存資訊【請依寄存機構、日期、號碼順序註記】

國外寄存資訊【請依寄存國家、機構、日期、號碼順序註記】

【序列表】(請換頁單獨記載)

申請專利範圍

- 1、 一種文件管理系統，係取得記錄於複數台電腦或伺服器的數位資訊，
分析該所取得的數位資訊，其具備：
執行緒分類部，其係確認該數位資訊中所含的文件資料各個的附帶資訊；基於該附帶資訊而將該文件資料分類為執行緒；
類似度解析部，其係將所分類的文件資料的附帶資訊中所含的要素擷取為各該執行緒，基於該所擷取的要素而解析該執行緒間的類似度；
及
統一部，其係基於該類似度而統一該執行緒彼此。
- 2、 如申請專利範圍第 1 項之文件管理系統，其中進一步具備：文件資料分類部，其係將該數位資訊中所含的文件資料分類為單獨文件資料、主文件資料、和該主文件資料有從屬關係的從屬文件資料、及其他的文件資料，
該執行緒分類部至少將該單獨文件資料及該主文件資料進行分類。
- 3、 如申請專利範圍第 1 或 2 項之文件管理系統，其中該其他的文件資料包含沒有附帶資訊的非構造化文件資料，
該執行緒分類部將該非構造化文件資料進行分類之際，參考該非構造化文件資料的內容。
- 4、 如申請專利範圍第 1 至 3 項中任一項之文件管理系統，其中該執行緒分類部處理該文件資料的標題、作成日期和時間、更新日期和時間、作者名之中至少任一個作為附帶資訊。

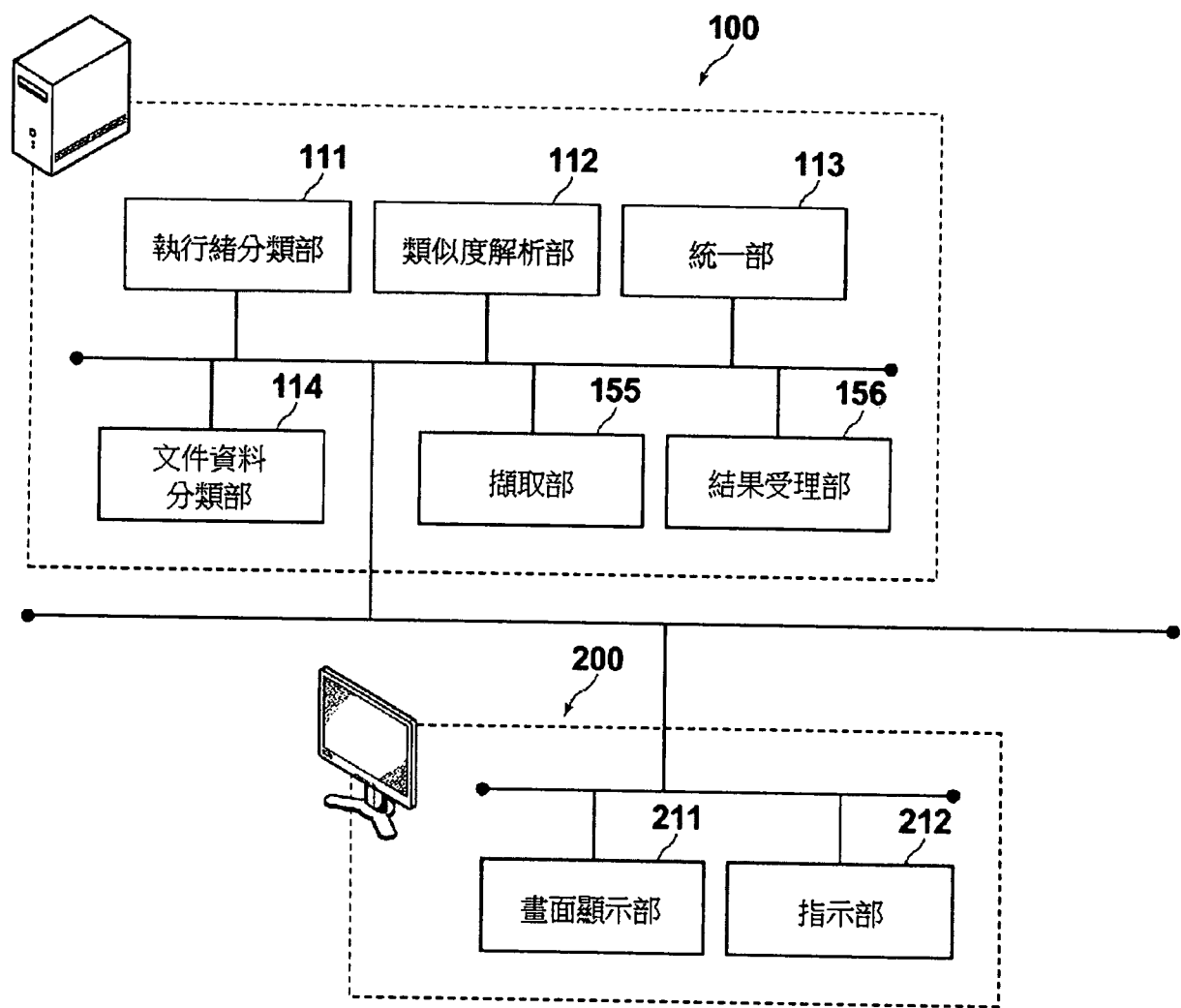
- 5、 如申請專利範圍第 1 至 4 項中任一項之文件管理系統，其中該文件管理系統進一步具備：
擷取部，其係從該所統一的執行緒擷取預定數量的文件資料；
畫面顯示部，其係在畫面上顯示該所擷取的文件資料；
結果受理部，其係受理利用者對於該所顯示的文件資料進行的相關性判斷結果；
選定部，其係基於該判斷結果，將該所擷取的文件資料分類為各判斷結果，在該所分類的文件資料中解析、選定共通出現的要素；
要素記錄部，其係記錄該選定的要素；
檢索部，其係從該文件資料檢索記錄於該要素記錄部的要素；及
分數計算部，其係使用該檢索部的檢索結果與該選定部的解析結果，計算表示該判斷結果與文件資料相關性的分數。
- 6、 如申請專利範圍第 2 項之文件管理系統，其中該統一部將該從屬文件資料之各個文件資料統一為和對應的主文件資料相同的執行緒。
- 7、 如申請專利範圍第 2 項之文件管理系統，其中該執行緒分類部進一步將該從屬文件資料之各個文件資料分類為和對應的主文件資料相同的執行緒。
- 8、 一種文件管理方法，係取得記錄於複數台電腦或伺服器的數位資訊，分析該所取得的數位資訊，電腦執行以下步驟：
確認該數位資訊中所含的文件資料各個的附帶資訊，基於該附帶資訊而將該文件資料分類為執行緒；
將所分類的文件資料的附帶資訊中所含的要素擷取為各該執行緒，基

於該所擷取的要素而解析該執行緒間的類似度；及

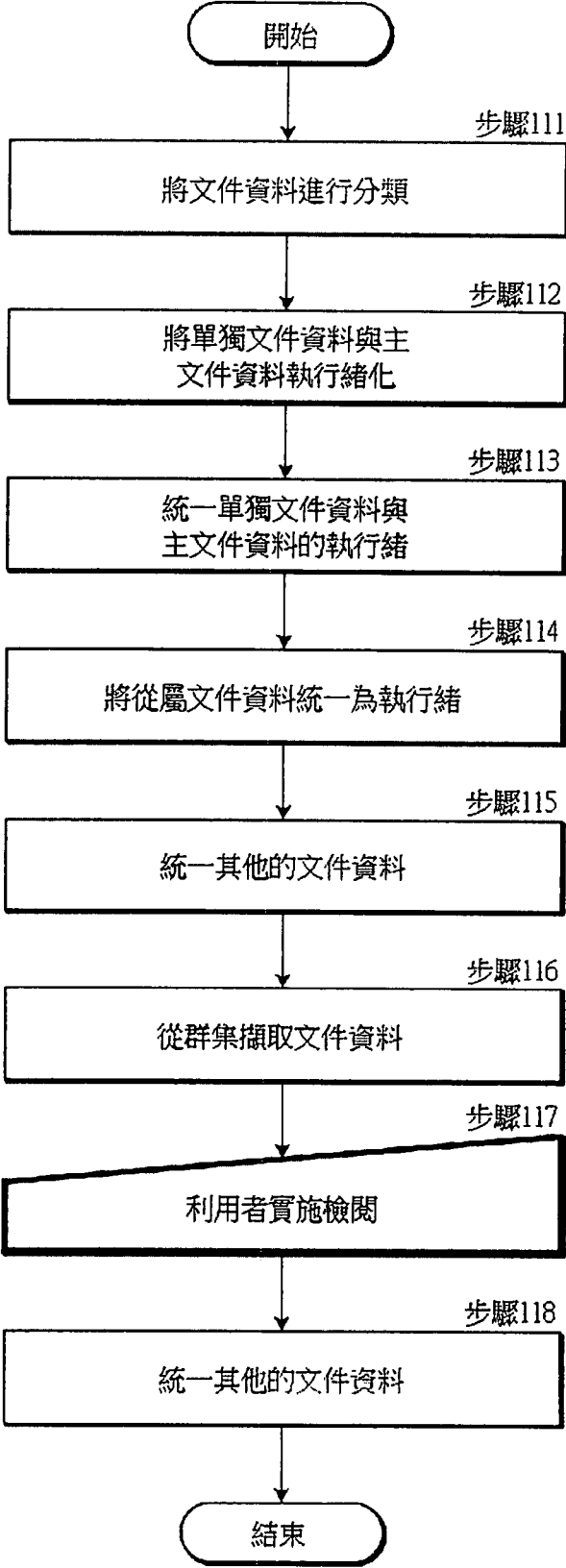
基於該類似度而統一該執行緒彼此。

- 9、 一種文件管理程式，係取得記錄於複數台電腦或伺服器的數位資訊，
分析該所取得的數位資訊，使電腦實現以下功能：
確認該數位資訊中所含的文件資料各個的附帶資訊，基於該附帶資訊
而將該文件資料分類為執行緒；
將所分類的文件資料的附帶資訊中所含的要素擷取為各該執行緒，基
於該所擷取的要素而解析該執行緒間的類似度；及
基於該類似度而統一該執行緒彼此。

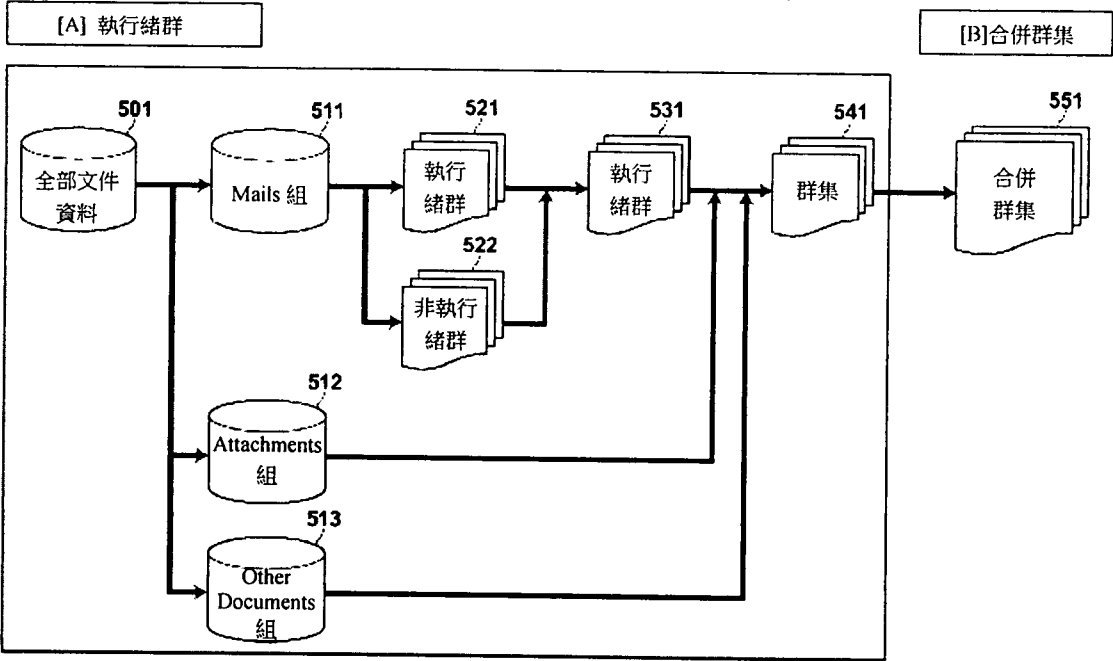
圖式



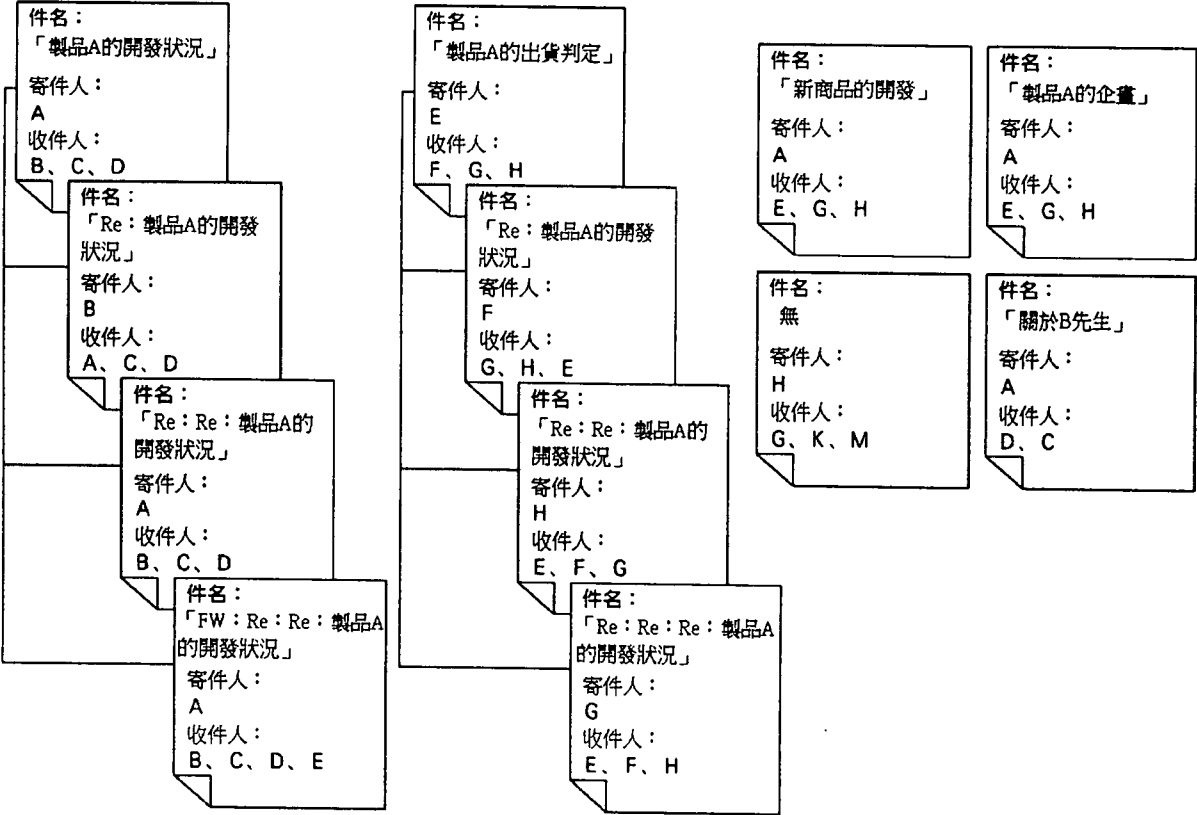
第 1 圖



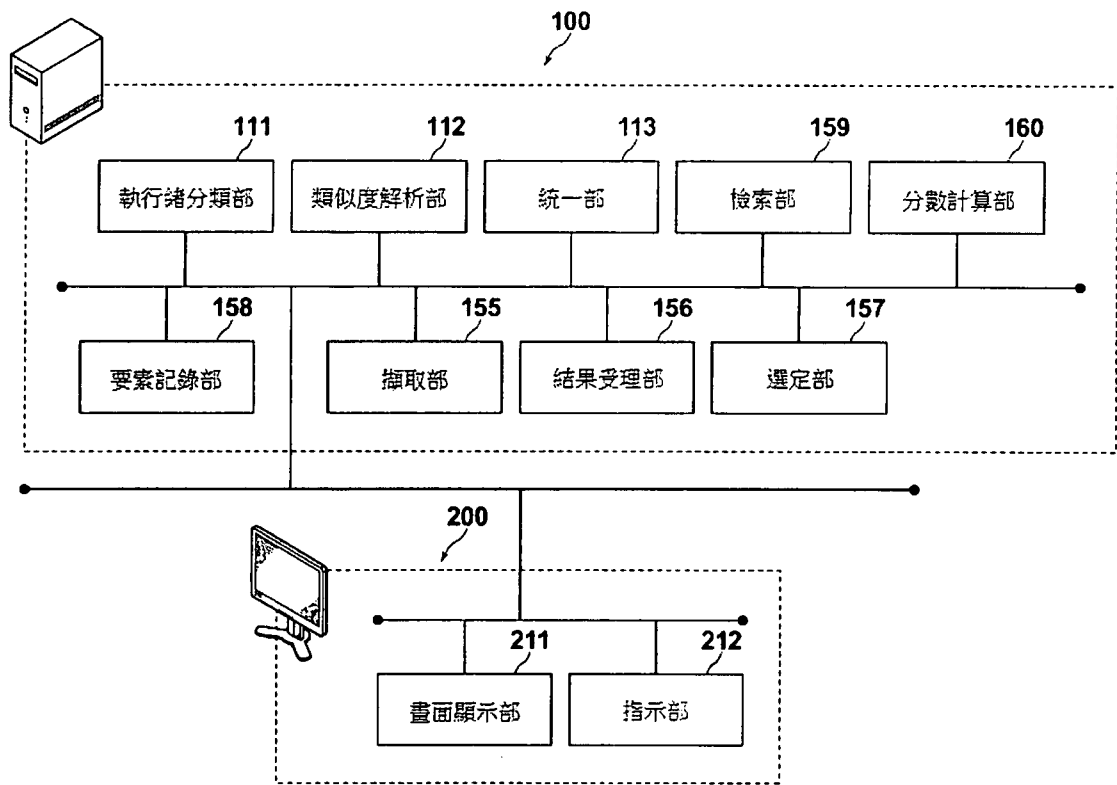
第 2 圖



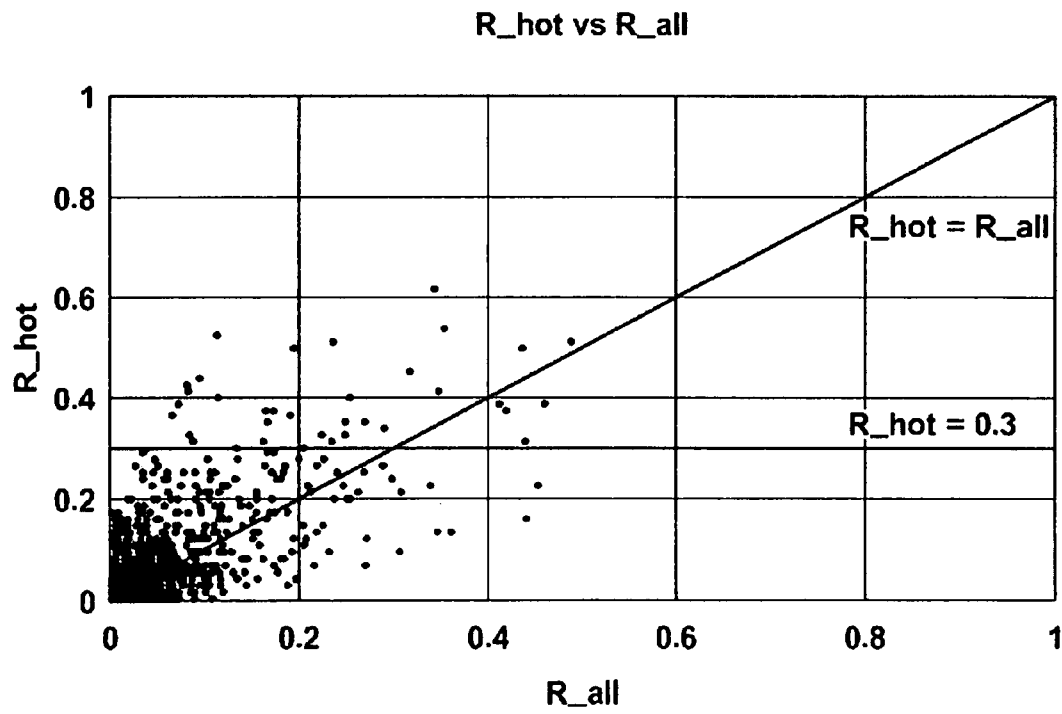
第 3 圖



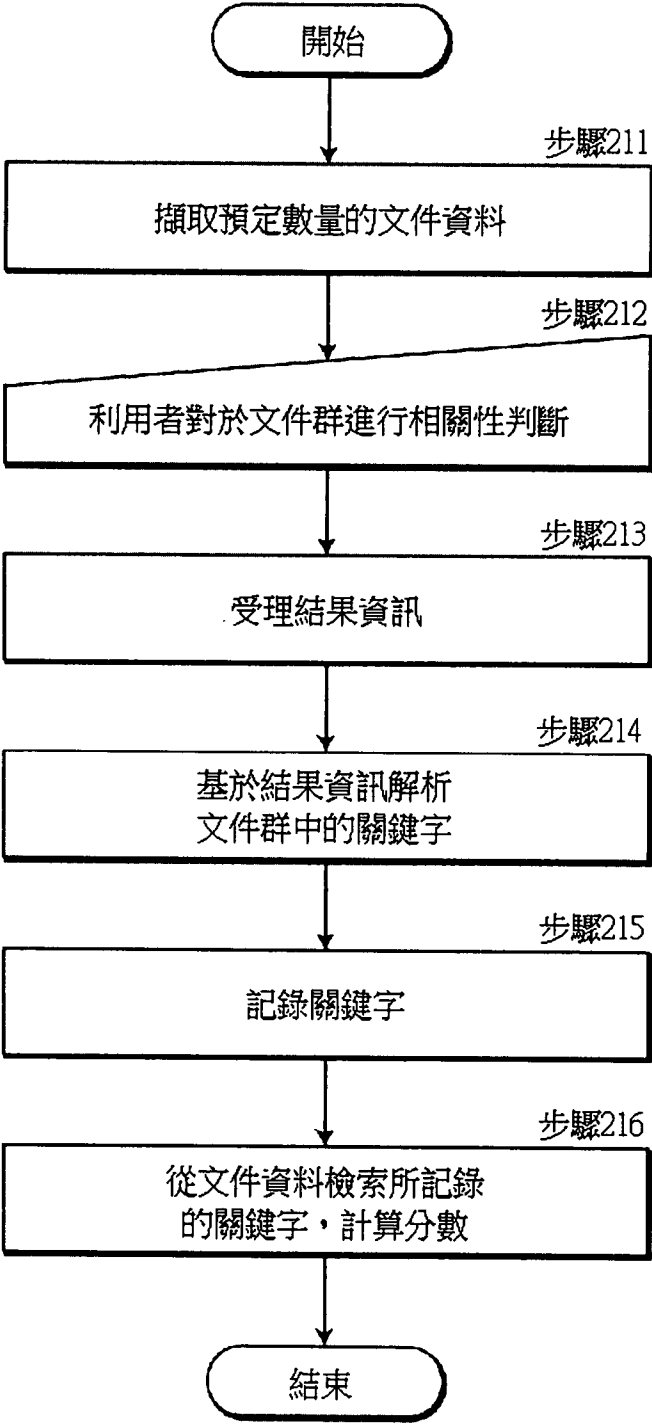
第 4 圖



第 5 圖



第 6 圖



第 7 圖