

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※ 申請案號： 97129953

※ 申請日期： 97.8.6

※IPC 分類： 606K 9/04 (2006.01)

606F 17/40 (2006.01)

606F 17/30 (2006.01)

606F 17/01 (2006.01)

一、發明名稱：(中文/英文)

具有辨識分類搜尋儲存功能的文件控管系統及方法

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

眾來科技股份有限公司

代表人：(中文/英文) 林益邦

住居所或營業所地址：(中文/英文)

台北市濟南路二段 29-1 號

國 籍：(中文/英文) 中華民國/ROC

三、發明人：(共 4 人)

姓 名：(中文/英文)

1. 劉立恩

2. 林益邦

3. 陳衍彰

4. 林聖富

國 籍：(中文/英文) 1.-4. 中華民國/ROC

四、聲明事項：

☐ 主張專利法第二十二條第二項☐第一款或☐第二款規定之事實，其事實發生日期為： 年 月 日。

☐ 申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

☐ 有主張專利法第二十七條第一項國際優先權：

☐ 無主張專利法第二十七條第一項國際優先權：

☐ 主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

☐ 主張專利法第三十條生物材料：

☐ 須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

☐ 不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

本發明係關於一種文件控管裝置及文件控管方法，尤指一種具有辨識分類搜尋儲存功能的文件控管裝置及文件控管方法。

【先前技術】

傳統的文件控管裝置，例如 TW-200500899(相當於 US-20040267557, CN-1567326)，雖然可以將用戶傳輸的電子文件，放置於與此電子文件所指定之位址相對應之資料夾中，但因依方法儲存的電子文件，日後要搜尋該檔案時，只有靠記憶找到該資料夾，再從該資料夾的大量檔案中，逐筆搜尋出所需要的電子文件，這對用戶造成很大的困擾。本發明利用光學識別器、特徵標記識別器等技術，在存檔同時，自動建立特徵標記索引，讓日後用戶只要隨便鍵入該電子文件的任一或多個特徵標記，就能馬上找到該電子文件。

【發明內容】

本發明之一目的，在提供一種文件控管裝置。

本發明之另一目的，在提供一種具有辨識、分類、搜尋、儲存功能的文件控管裝置。

本發明之又一目的，在提供一種藉由光學字符識別器辨讀特徵標記的文件控管裝置。

本發明之再一目的，在提供一種以特徵標記做為文件索引的文件控管裝置。

本發明之另一目的，在提供一種藉由光學字符識別器辨讀特徵標記，並以該特徵標記做為文件索引的文件控管裝置。

本發明之又一目的，在提供一種藉特徵標記搜尋文件再經由網頁伺服器輸出的文件控管裝置。

本發明之再一目的，在提供一種包含網頁伺服器、檔案接收裝置、光學字符識別器和資料庫的文件控管裝置。

本發明之又一目的，在提供一種具有文件控管功能的多功事務裝置。

本發明之另一目的，在提供一種具有辨識、分類、搜尋、儲存功能的多功事務裝置。

本發明之又一目的，在提供一種藉由光學字符識別器辨讀特徵標記的多功事務裝置。

本發明之再一目的，在提供一種以特徵標記做為文件索引的多功事務裝置。

本發明之另一目的，在提供一種藉由光學字符識別器辨讀特徵標記，並以該特徵標記做為文件索引的多功事務裝置。

本發明之又一目的，在提供一種藉特徵標記搜尋文件再經由網頁伺服器輸出的多功事務裝置。

本發明之再一目的，在提供一種包含網頁伺服器、檔案接收裝置、光學字符識別器和資料庫的多功事務裝置。

本發明之另一目的，在提供一種具有辨識、分類、搜尋、儲存功能的文件控管方法。

本發明之又一目的，在提供一種藉由光學字符識別器辨讀特徵標記的文件控管方法。

本發明之再一目的，在提供一種以特徵標記做為文件索引的文件控管方法。

本發明之另一目的，在提供一種藉由光學字符識別器辨讀特徵標記，並以該特徵標記做為文件索引的文件控管方法。

本發明之又一目的，在提供一種藉特徵標記搜尋文件再經由網頁伺服器輸出的文件控管方法。

本發明之再一目的，在提供一種包含網頁伺服器、檔案接收裝置、光學字符識別器和資料庫的文件控管方法。

本發明為一種具有辨識分類搜尋儲存功能的文件控管裝置，其包含：

一檔案接收裝置，用以讀取文件；

一光學識別器，用以對該檔案接收裝置讀取的文件中的非文字內容進行光學識別；

一特徵標記識別器，用以建立該輸入文件的特徵標記；及

一資料庫，用以儲存該讀入文件，及/或依需要將資料庫的文件輸出或經由該檔案接收裝置輸出；

其特徵在：

該光學識別器可自動對該輸入文件的非文字部分進行光學辨識，得到光學辨識結果；

該特徵標記識別器，用以依該文件的特徵內容，建立該文件的特徵標記，其中該文件的特徵內容包含該文件的文字內容及/或該光學辨識的結果；

其儲存文件時，係依該檔案接收裝置讀入的來源辨識資訊及/或該文件的特徵標記進行分類，做為儲存該輸入資料的依據；及

其於儲存該文件時，依該特徵標記建立索引，用以做為裝置欲輸出文件時，搜尋該筆文件的依據。

上述文件控管裝置，可單獨存在，亦可配合多功能事務機(以下簡稱 MFP，例如掃描機、傳真機、影印機等)使用，一般而言，以配合多功能事務機(以下簡稱 MFP)使用為較佳。

上述檔案接收裝置可為任意習知的檔案接收裝置，用以透過匯流排(Bus)及/或數據傳輸協定 (USB / RS232 等)接收來自上述 MFP 的電子文件；或接收來自外部網頁伺服器或外部實體傳輸線的電子文件。當然該檔案接收裝置也可以是自己完成、委外完成或合作完成的類似檔案接收裝置。

上述所謂文件，係指可藉由 MFP 轉換成電子文件的文件，或電子文件本身(例如經由電子郵件的本文及/或附件、傳真機傳送的電子檔、掃描裝置讀入的電子檔、電腦所產生的各式電子檔案、來自外部網頁伺服器或外部實體傳輸線的電子文件等)；或經由轉換技術所得到的電子資訊，例如紙本文件(文字資料、圖式、表單等)、照片等可透過掃

瞄裝置轉成電子文件；實物、樣品等可透過數位照相轉成電子文件；或任意可轉換成電子文件的資訊所轉成的電子資訊。文件格式並無一定限制，例如 TXT、MS-Office、PDF、JPG、GIF、TIFF、HTML 等。

上述該來源辨識資訊，其可為任意可辨識文件來源的資訊，諸如檔頭資訊，例如傳送者、傳送者帳號、主旨、傳送來源（例如來自外部網路伺服器的主機名稱、MAC 位址、網路位址/IP Address 等）、檔案名稱、傳送日期、檔案格式、檔案內容摘要等。

上述光學識別器，可為任意習知的光學識別器，諸如光學字符識別器（例如 ABBYY 公司的 FINE READER）、條碼識別器（例如一般的一維條碼識別器、二維條碼識別器）等；當然也可以是自己完成、委外完成或合作完成的類似光學識別器。若該光學識別器為條碼識別器，則客戶必須使用條碼，造成客戶的困擾，因此，一般而言以使用光學字符識別器為較佳。

若該讀入文件只含文字內容時，則該文字內容即為該文件的特徵內容。

若該讀入文件不含文字內容時，則該光學識別器之辨識結果即為該文件的特徵內容。

若該讀入文件同時含文字內容時和非文字內容時，其可以單純光學識別器之辨識結果、或單純文字內容、或文字內容加辨識結果做為該文件的特徵內容。一般而言，當光學識別器為光學字符識別器時，通常以文字內容加辨識

結果做為該文件的特徵內容；當光學識別器為條碼識別器時，通常以辨識結果做為該文件的特徵內容。

上述特徵標記識別器，可採用任意習知的特徵標記識別器，例如意藍公司的龍捲風搜尋引擎；當然也可以是自己完成、委外完成或合作完成的類似特徵標記識別器。

上述之特徵標記識別器，其對該文件的特徵內容，進行諸如斷詞斷句、關鍵詞擷取及/或文件內涵分析，以建立該文件的特徵標記。一般而言，該特徵標記識別器除上述功能外，以進一步具有新詞學習、用字/用語/詞性/意境分析等功能為較佳。

在特殊狀況下，例如：經特徵標記識別器辨識後，不含特徵標記時，必要時，裝置可要求使用者提示特徵標記，或自動對該特徵標記加註諸如”其他類別”做為特徵標記。此外，該特殊狀況，必要時，可列入諸如新詞學習、意境分析的統計/分析/資料探勘(data mining)等後續程序。

上述資料庫於儲存文件時，可依該檔案接收裝置讀入的來源辨識資訊(例如檔頭)進行分類，儲存該輸入資料，例如分類方式(資料夾)為：

<A001 公司> (外部文件來源 1)

<A002 公司> (外部文件來源 2)

.....

<B001 部門> (內部文件來源 1)

<B002 部門> (內部文件來源 2)

.....

其中 A001 公司、A002 公司等，可以是各該公司的公司名稱、公司代號、公司網域名稱、公司電話號碼等，及/或其組合，該等文件係經由傳真機所輸入，或透過外部網路伺服器輸入，經由檔案接收裝置讀入的文件，因此列為外部文件；B001 部門、B002 部門等，可以是各該部門的部門名稱、部門代號、部門網域名稱、部門分機號碼等，及/或其組合，該等文件係透過內部匯流排及/或傳輸硬體輸入，經由檔案接收裝置讀入的文件，因此列為內部文件。

上述資料庫於儲存文件時，也可依該檔案接收裝置讀入的來源辨識資訊(例如檔頭)進行分類及進一步分類，以儲存該該輸入資料，例如分類方式(資料夾)為：

<A001 公司> (外部文件來源 1)

<A001 公司 -001>

<A001 公司 -002>

<A001 公司 -003>

.....

<A002 公司> (外部文件來源 2)

<A002 公司 -001>

<A002 公司 -002>

<A002 公司 -003>

.....

.....
<B001 部門> (內部文件來源 1)

<B001 部門 -001>

<B001 部門-002>

<B001 部門-003>

.....

<B002 部門> (內部文件來源 2)

<B002 部門-001>

<B002 部門-002>

<B002 部門-003>

.....

.....

其中 A001 公司、A002 公司等，可以是各該公司的公司名稱、公司代號、公司網域名稱、公司電話號碼等，及/或其組合。A001 公司-001、A001 公司-002、A001 公司-003 等分別為 A001 公司的部門別或部門代號、使用者名稱(檔頭資訊為電子郵件位址時)、或該公司自訂的分類方式；A002 公司-001、A002 公司-002、A002 公司-003 等分別為 A002 公司的部門別或部門代號、使用者名稱(檔頭資訊為電子郵件位址時)、或該公司自訂的分類方式；因此該分類方式可以多於兩層的分類。

B001 部門、B002 部門等，可以是各該部門的部門名稱、部門代號、部門網域名稱、部門分機號碼等，及/或其組合。B001 部門-001、B001 部門-002、B001 部門-003 等分別為 B001 部門的次單位名稱別或次單位代號、使用者名稱(檔頭資訊為電子郵件位址時)、或該部門自訂的分類方式；B002 部門-001、B002 部門-002、B002 部門-003 等分

別為 B002 部門的次單位名稱別或次單位代號、使用者名稱(檔頭資訊為電子郵件位址時)、或該部門自訂的分類方式；因此該分類方式可以多於兩層的分類。

必要時，上述分類亦可納入上述特徵內容或特徵標記，做為分類的依據之一，但以不納入該等資訊做為分類的依據為較佳。

上述資料庫於儲存文件時，可依該檔案接收裝置讀入的來源辨識資訊(例如檔頭)、特徵內容、特徵標記、儲存日期時間、及/或流水號等，做為儲存檔案時的檔名，例如 A001 公司的檔案儲存為：

<A001 公司> (外部文件來源 1)

BX001-a1 說明.doc (檔名 1)

BX002-a1 規範.xls (檔名 2)

BX003-a2 內容.doc (檔名 3)

BX004-a3 介紹.pdf (檔名 4)

.....

其中 BX001、BX002、BX003、BX004 為流水號，"a1 說明.doc"、"a1 規範.xls"、"a2 內容.doc"、"a3 介紹.pdf"的主檔名係裝置根據部份特徵內容自動訂定的，而副檔名係依據各該檔案格式自動訂定的。

若主檔名都含流水號，則各分類(含細分類)內的檔案都不會重複。但檔名不含流水號時，在特殊狀況下，新文件自動產生的檔名，可能和相同分類(含細分類)內的既存文件的檔名相同，此時，裝置可要求使用者提示新的檔名，

或自動對該特徵標記加註諸如日期(及/或時間)等辨識碼。
特殊狀態下，諸如檔名不具特異性，例如主檔名為空符號或資料庫禁制符號時，此時，裝置可要求使用者提示新的檔名，或自動對該特徵標記加註諸如日期(及/或時間)等辨識碼。

上述特徵標記，其可為一個或複數個特徵用字及/或特徵用語的集合。當建立索引時，以各單一特徵用字或特徵用語分別建立該文件的索引為主，但亦可進一步建立以複數個特徵用字及/或特徵用語的文件索引，但一般在檢索時採用"and"功能取代後者。例如檔案 1 經光學識別器辨識後，得到的特徵內容為："...XX1...XX2...XX3XX4..."，經特徵識別器辨識後，得到的特徵用語為：XX1、XX2、XX3、XX4、XX3XX4.....等，其中特徵用語 XX3XX4 為特徵用語 XX3 和特徵用語 XX4 的複合特徵用語，且裝置自動設定檔名為"YYY"；而檔案 2 經光學識別器辨識後，得到的特徵內容為："...XX1...XX3...XX4...XX5..."，經特徵識別器辨識後，得到的特徵用語為：XX1、XX3、XX4、XX5.....等，且裝置自動設定檔名為"ZZZ"；則裝置將自動產生特徵用語索引如下：

XX1.....YYY

XX1.....ZZZ

XX2.....YYY

XX3.....YYY

XX3.....ZZZ

XX3XX4.....YYY

XX4.....YYY

XX5.....ZZZ

當使用者想瀏覽或輸出其所儲存的文件時，可憑使用者名稱(或代號、電話號碼、分機號碼等)，必要時，配合密碼(例如文字密碼、條碼、指紋、虹膜等)，檢索待瀏覽或輸出的文件，檢索方式含任意習知的檢索方式，例如全文檢索、關鍵字(特徵用語、特徵用詞)檢索、分類檢索、日期及/或時間檢索，或日期區間等檢索。以特徵用語檢索為例，例如上述案例中：

日後該使用者想要檢索含 XX1 的檔案，可找到檔案 YYY 和檔案 ZZZ(當然可能還有其他也含 XX1 的檔案)；

日後該使用者想要檢索含 XX2 的檔案，可找到檔案 YYY，但無法找到檔案 ZZZ；

日後該使用者想要檢索同時含 XX3 和 XX4 的檔案，即可找到檔案 YYY 和檔案 ZZZ；

日後該使用者想要檢索含 XX3XX4 的檔案，只能找到檔案 YYY，無法找到檔案 ZZZ。

上述文件控管裝置可以配合多功能事務機使用一如上述，但一般而言，亦可和 MFP 結合形成具有文件控管功能的多功事務裝置，如下述。

本發明也包括一種具有文件控管功能的多功事務裝置，其包含：

一事務部件；

- 一 檔案接收裝置，用以由該事務部件讀取文件；
- 一 光學識別器，用以對該檔案接收裝置讀取的文件中的非文字內容進行光學識別；
- 一 特徵標記識別器，用以建立該輸入文件的特徵標記；及
- 一 資料庫，用以儲存該讀入文件，及/或依需要將資料庫的文件輸出、經由該檔案接收裝置輸出，或經由該該事務部件輸出；

其特徵在：

- 該光學識別器可自動對該輸入文件的非文字部分進行光學辨識，得到光學辨識結果；
- 該特徵標記識別器，用以依該文件的特徵內容，建立該文件的特徵標記，其中該文件的特徵內容包含該文件的文字內容及/或該光學辨識的結果；
- 其儲存文件時，係依該檔案接收裝置讀入的來源辨識資訊及/或該文件的特徵標記進行分類，做為儲存該輸入資料的依據；及
- 其於儲存該文件時，依該特徵標記建立索引，用以做為裝置欲輸出文件時，搜尋該筆文件的依據。

上述所謂電子文件、來源辨識資訊、分類方式、檔案儲存方式、特徵標記辨識方式、建立索引方式、輸出電子文件方式，都類如前述。

上述檔案接收裝置、光學識別器、光學識別器、資料庫，都類如前述。

上述事務部件可為任意具有文件輸出及/或輸入功能的

習知裝置(例如傳統的印表機、掃描機、傳真機、影印機)，或類似的裝置，用以實際上連接上述檔案接收裝置、光學識別器、光學識別器、資料庫，構成一完整的具有文件控管功能的多功事務裝置；或由該等具有文件輸出及或輸入功能的習知裝置(例如傳統的印表機、掃描機、傳真機、影印機)加以修正或調整，用以實際上連接上述檔案接收裝置、光學識別器、光學識別器、資料庫，構成一完整的具有文件控管功能的多功事務裝置；或依諸如傳統的印表機、掃描機、傳真機、影印機機制，設計具有文件輸出及或輸入功能的部件，用以實際上連接上述檔案接收裝置、光學識別器、光學識別器、資料庫，構成一完整的具有文件控管功能的多功事務裝置。

本發明之文件控管方法，其包括：

一文件接收步驟，用以接收待儲存的電子文件；

一文件分解步驟，用以分解該電子文件的來源辨識資訊；

一分類步驟，用以依該來源辨識資訊進行分類；及

一檔案儲存步驟，用以依該分類儲存該電子文件；

其特徵在其進一步包括：

一特徵標記辨識步驟，用以該電子文件的內容辨識特徵標記；及

一索引建立步驟，用以依該特徵標記建立索引，作為裝置欲輸出該電子文件時，搜尋該筆該電子文件的依據。

上述所謂電子文件、來源辨識資訊、分類方式、檔案

儲存方式、特徵標記辨識方式、建立索引方式、輸出電子文件方式，都類如前述。本發明方法的實際處理程序則如下述較佳具體例或實施例所述。

上述方法中，若分類係採用前述依來源辨識資訊進行粗分類，再依特徵標記進行細部分類，則該分類步驟和特徵標記辨識步驟的關係，可以是：依來源辨識資訊進行粗分類，而後等進行特徵標記辨識步驟後，再依特徵標記進行細部分類；也可以是：先執行特徵標記辨識步驟，再執行分類步驟(含粗分類和細部分類)。

上述方法中，若分類係單純依來源辨識資訊進行分類，不再依特徵標記進行細部分類，則該分類步驟和特徵標記辨識步驟的關係，可以是：先執行分類步驟，再執行特徵標記辨識步驟，這種情況下，特徵標記辨識步驟和檔案儲存步驟其先後順序也可以對調，甚至是實質上同步進行或交互進行；也可以是：先執行特徵標記辨識步驟，再執行分類步驟；當然兩者實質上同步進行或交互進行也可以。

【實施方式】

為進一步說明本發明，茲以較佳具體例配合圖式說明於後：

圖 1a 中，外部(傳真發送單位 300)發送傳真(310)給具有文件控管裝置(100)的 MFP 時，該文件控管裝置(100)於接收到傳真(180)後，就執行存檔任務(190)。

圖 1b 中，內部使用具有文件控管裝置(100)的 MFP(200)儲存文件時，首先利用 MFP(200)掃描該文件，並傳送到文件控管裝置(100)，該文件控管裝置(100)於接收到傳真(180)後，就執行存檔任務(190)。

圖 1c 中，內部使用電腦(205)傳送文件(210)給文件控管裝置(100)，當文件控管裝置(100)於接收到傳真(180)時，就執行存檔任務(190)。

圖 2 中，當使用者欲檢索既有電子文件時，可透過電腦(205)上傳一或多個特徵用語(292)至文件控管裝置；當文件控管裝置接收到該上傳資料(192)，立即執行檢索任務(194)，而後將檢索結果(傳送合乎檢索條件的檔案，或下傳”無”的訊息)下傳給電腦(196)。

圖 3a 為本發明文件控管裝置配合 MFP 使用的結構方塊圖，其中 100 為文件控管裝置，120、130、140、150 分別為文件控管裝置的檔案接收裝置、光學字符識別器(OCR)、資料庫和特徵標記識別器；200 為配合該文件控管裝置的 MFP，210、220、230、240 分別為 MFP 的傳真機制、掃描機制、印表機制、影印機制。其中該文件控管裝置 100，和該 MFP(200)係以一傳輸部件連接，圖中以一線段(未標號)表示。本裝置操作模式如前述。

圖 3b 為本發明文件控管裝置和 MFP 結合成一體的結構方塊圖，其中 100 為文件控管裝置，120、130、140、150 分別為文件控管裝置的檔案接收裝置、光學字符識別器(OCR)、資料庫和特徵標記識別器；200 為和該文件控管裝

置構成亦體的 MFP，210、220、230、240 分別為 MFP 的傳真機制、掃描機制、印表機制、影印機制。其中該文件控管裝置 100，和該 MFP(200)已結合成一體，圖中以一外框(未標號)表示。本圖也可以視同具有文件控管功能的多功事務裝置的結構方塊圖。本多功事務裝置操作模式如前述。

圖 3c 為本發明具有文件控管功能的多功事務裝置的結構方塊圖，其中 800 為多功事務裝置，810、820、830、840、850、860、870、880 分別為多功事務裝置的檔案接收裝置、光學字符識別器(OCR)、資料庫和特徵標記識別器、傳真機制、掃描機制、印表機制、影印機制。本多功事務裝置操作模式如前述。

圖 4 顯示：當文件控管裝置接收電子文件(510)後，立即分解該電子文件的檔頭(520)，依檔頭粗分類(530)，並將電子文件中的非文字進行光學字符識別(540)，而後，依 OCR 辨識結果，配合該電子文件的文字內容，建立該文件之特徵內容(550)，再利用龍捲風搜尋引擎，由特徵內容(550)辨識特徵標記(560)；而後，一方面依特徵標記建立索引(570)，作為裝置欲輸出文件時，搜尋該筆文件的依據；另一方面依特徵標記，進行細部分類(580)，而後依分類結果(粗分類加細分類)，儲存該電子文件(590)。

圖 5 顯示：當文件控管裝置接收電子文件(510)後，立即分解該電子文件的檔頭(520)，並依檔頭進行分類(530)，而後，將電子文件中的非文字進行光學字符識別(540)，並

依 OCR 辨識結果，依序建立該文件之特徵內容(550)，再依特徵內容辨識特徵標記(560)，並依特徵標記建立索引(570)，作為裝置欲輸出文件時，搜尋該筆文件的依據，最後依檔頭分類儲存該電子文件(590)。

圖 6 顯示：當文件控管裝置接收電子文件(510)後，立即分解該電子文件的檔頭(520)，依檔頭進行分類(530)，並依檔頭分類儲存該電子文件(590)；而後，將電子文件中的非文字進行光學字符識別(540)；將 OCR 的結果和文件中的文字內容合併為特徵內容(550)，再依特徵內容辨識特徵標記(560)，並依特徵標記建立索引(570)，作為裝置欲輸出文件時，搜尋該筆文件的依據。

圖 7 為圖 4 中步驟 560 加上細部分類步驟(580)的細部流程圖，其係在特徵標記中找關鍵字，看看是否含關鍵字(581)，若含關鍵字依關鍵字進一步分類(582)，以完成細分類(586)；若不含關鍵字，則由使用者決定是否手動分類(至 583)，若是則依其鍵入內容做為細部分類(584)，以完成細分類(546)；若否，則無細部分類(585)，亦即完成細分類(586)。

圖 8 顯示細部分類步驟(580)的另一細部流程圖，其係在特徵標記中找關鍵字，看看是否含關鍵字(581)，若含關鍵字依關鍵字進一步分類(582)，以完成細分類(586)；若不含關鍵字，則無細部分類(585)，亦即完成細分類(586)。

圖 9 顯示：裝置在接收到檢索資訊(步驟 610)後，立即依檢索條件進行檢索(步驟 620)，判斷是否有合乎條件的檔

案(步驟 630)，若有合乎條件的檔案，就將合乎條件的檔案傳送給使用者(步驟 640)；若沒有合乎條件的檔案，就傳送“沒有合乎條件的檔案”的訊息給使用者(步驟 650)。

圖 10 為本發明方法一較佳具體例的流程圖。當文件控管裝置接收電子文件(510)後，立即分解該電子文件的檔頭(520)，依檔頭進行分類(530)，並依檔頭分類儲存該電子文件(590)；而後，判斷文件中是否含“非文字內容”(542)，若含“非文字內容”，則將電子文件中的非文字進行光學字符識別(540)，而後進行步驟 550(將 OCR 的結果和文件中的文字內容合併為特徵內容)；若不含“非文字內容”，則直接進行步驟 550(直接以文件中的文字內容做為特徵內容)；而後，依特徵內容辨識特徵標記(560)，並依特徵標記建立索引(570)，作為裝置欲輸出文件時，搜尋該筆文件的依據。

另以本案說明書為例，說明本案步驟 540、步驟 550、步驟 560 和步驟 570，以及檢索該檔案的狀況如下：

本案內容包括：發明名稱、發明摘要、發明說明、申請專利範圍、圖式……等，其中發明名稱、發明摘要、發明說明、申請專利範圍……等為文字內容，圖式為非文字內容，因此在步驟 540 中，光學字符識別器將對圖式進行光學字符辨識。以圖 1a 為例，OCR 後，會得到“300 傳真發送單位”、“310 傳真文件”、“100 文件控管管裝置”、“180 接收文件”、“190 執行存檔任務”等文字內容。

步驟 550 中，會將步驟 540 辨識所得的文字內容，和原來的文字內容(發明名稱、發明摘要、發明說明、申請專

利範圍.....等)合併，成為特徵內容。

步驟 560 中，特徵標記辨識器會對步驟 550 所建立的特徵內容，進行特徵標記辨識。以發明名稱進行特徵標記辨識為例，將會得到”辨識、分類、搜尋、儲存、文件、控管、裝置”等特徵用語，以圖 1 經 OCR 後所得的文字進行特徵標記辨識為例，將會得到”傳真、發送、單位、文件、控管、裝置、接收、執行、存檔、任務”等特徵用語。

在步驟 570 中，裝置會依步驟 560 辨識所得的特徵用語，對待存檔案(圖 3 或圖 4 所示的流程)或既存檔案(圖 5 所示的流程)，進行建立索引的程序。設若裝置自動設定檔名為”具有辨識分類搜尋儲存功能的文件控管裝置”(下文中簡化為文件控管裝置)，以發明名稱內所含的特徵用語為例，裝置將自動產生特徵用語索引如表 1：

表 1：以發明名稱內的特徵用語所建立的索引表

辨識	文件控管裝置
辨識分類	文件控管裝置
辨識分類搜尋	文件控管裝置
辨識分類搜尋儲存	文件控管裝置
辨識分類搜尋儲存功能	文件控管裝置
分類	文件控管裝置
分類搜尋	文件控管裝置
分類搜尋儲存	文件控管裝置
分類搜尋儲存功能	文件控管裝置
搜尋	文件控管裝置

搜尋儲存

搜尋儲存功能

儲存

儲存功能

文件

文件控管

文件控管裝置

控管

控管裝置

裝置

文件控管裝置

文件控管裝置

文件控管裝置

文件控管裝置

文件控管裝置

文件控管裝置

文件控管裝置

文件控管裝置

文件控管裝置

文件控管裝置

再以圖 1 所含的特徵用語為例，裝置將自動產生新增的特徵用語索引如表 2：

表 2：以圖 1 中的特徵用語所建立的索引表

傳真

文件控管裝置

傳真發送

文件控管裝置

傳真發送單位

文件控管裝置

發送

文件控管裝置

發送單位

文件控管裝置

單位

文件控管裝置

傳真文件

文件控管裝置

接收

文件控管裝置

接收文件

文件控管裝置

執行

文件控管裝置

執行存檔	文件控管裝置
執行存檔任務	文件控管裝置
存檔	文件控管裝置
存檔任務	文件控管裝置
任務	文件控管裝置

表 2 中不含“文件、文件控管、文件控管裝置、控管、控管裝置、裝置”等特徵用語的索引，因為該等索引，在表 1 中已出現。

在存檔並建立索引表後，則裝置使用者可利用特徵用語檢索/顯示/下載該檔案。例如裝置使用者 100 以“儲存”進行特徵用語檢索，文件控管裝置 100 在步驟 610 中接收檢索資訊後，立即執行檢索任務 620，看看索引表中是否含“儲存”這個特徵用語(步驟 630)，檢索結果為索引表中含“儲存”這個特徵用語，因此接著執行步驟 640，亦即文件控管裝置會將本文件下載到使用者端。使用者接收該資訊後，可自行決定顯示及/或下載本文件。

【圖式簡單說明】

圖 1a 為本發明使用者接收傳真時的傳送/存檔機制示意圖。

圖 1b 為本發明使用者掃描文件並傳送資料的存檔機制示意圖。

圖 1c 為本發明使用者利用電腦傳送資料的存檔機制示

意圖。

圖 2 為本發明使用者檢索檔案，和文件控管裝置傳送檔案的機制示意圖。

圖 3a 為本發明文件控管裝置配合 MFP 使用的結構方塊圖。

圖 3b 為本發明文件控管裝置和 MFP 結合成一體的結構方塊圖。

圖 3c 為本發明具有文件控管功能的多功事務裝置的結構方塊圖。

圖 4 為本發明文件控管裝置執行儲存任務的一較佳執行流程示意圖。

圖 5 為本發明文件控管裝置執行儲存任務的另一較佳執行流程示意圖。

圖 6 為本發明文件控管裝置執行儲存任務的另一較佳執行流程示意圖。

圖 7 為本發明文件控管裝置中步驟 580(進行細部分類)的一較佳執行流程示意圖。

圖 8 為本發明文件控管裝置中步驟 580(進行細部分類)的另一較佳執行流程示意圖。

圖 9 為本發明文件控管裝置執行檢索任務的一較佳執行流程示意圖。

圖 10 為本發明文件控管方法的一較佳具體例的流程示意圖。

【主要元件符號說明】

- 100 文件控管裝置
 - 110 網頁伺服器
 - 120 檔案接收裝置
 - 130 光學字符識別器
 - 140 資料庫
 - 150 特徵標記識別器
- 200 會員的多功能事務機
 - 210 傳真機制
 - 220 掃描機制
 - 230 印表機制
 - 240 影印機制
 - 280 接收文件
 - 290 上傳文件
 - 292 上傳特徵用語
 - 296 接收下傳檔案
- 300 傳真發送單位
 - 310 傳真文件
- 500 存檔流程圖
 - 510 接收電子文件
 - 520 分解電子文件
 - 530 依檔頭粗分類
 - 540 光學字符辨識
 - 550 建立特徵內容
 - 560 辨識特徵標記

- 570 建立索引
- 580 進行細部分類
- 581 判斷有關鍵字否
- 582 依關鍵字分類
- 583 判斷手動分類否
- 584 依輸入值分類
- 585 無細部分類
- 586 完成(細部)分類
- 590 依分類存檔
- 600 檢索流程圖
 - 610 接收檢索資訊
 - 620 執行檢索任務
 - 630 判定有檔案否
 - 640 下傳檔案資訊
 - 650 下傳檢索結果
- 800 多功事務裝置
 - 810 檔案接收裝置
 - 820 光學字符識別器(OCR)
 - 830 資料庫
 - 840 特徵標記識別器
 - 850 傳真機制
 - 860 掃描機制
 - 870 印表機制
 - 880 影印機制

五、中文發明摘要：

一種具有辨識分類搜尋儲存功能的文件控管系統，其包含：

- 一網頁伺服器；
- 一檔案接收伺服器，用以藉由該網頁伺服器，讀取文件；
- 一光學識別器，用以對該檔案接收伺服器讀取的文件中的非文字內容進行光學識別；
- 一特徵標記識別器，用以建立該輸入文件的特徵標記；
- 及
- 一資料庫，用以儲存該讀入文件，及/或依需要將資料庫的文件經由該檔案接收伺服器和網頁伺服器輸出。

六、英文發明摘要：

十、申請專利範圍：

1. 一種具有辨識分類搜尋儲存功能的文件控管裝置，其包含：

一檔案接收裝置，用以讀取文件；

一光學識別器，用以對該檔案接收裝置讀取的文件中的非文字內容進行光學識別；

一特徵標記識別器，用以建立該輸入文件的特徵標記；
及

一資料庫，用以儲存該讀入文件，及/或依需要將資料庫的文件輸出或經由該檔案接收裝置輸出；

其特徵在：

該光學識別器可自動對該輸入文件的非文字部分進行光學辨識，得到光學辨識結果；

該特徵標記識別器，用以依該文件的特徵內容，建立該文件的特徵標記，其中該文件的特徵內容包含該文件的文字內容及/或該光學辨識的結果；

其儲存文件時，係依該檔案接收裝置讀入的來源辨識資訊及/或該文件的特徵標記進行分類，做為儲存該輸入資料的依據；及

其於儲存該文件時，依該特徵標記建立索引，用以做為裝置欲輸出文件時，搜尋該筆文件的依據。

2. 如申請專利範圍第 1 項所述之文件控管系統，其中該光學識別器為光學字符識別器。

3. 如申請專利範圍第 1 項所述之文件控管裝置，其中該裝置儲存文件時，係依該檔案接收裝置讀入的來源辨識資訊進行分類，做為儲存該輸入資料的依據。
4. 如申請專利範圍第 1 項所述之文件控管裝置，其中該來源辨識資訊為檔頭資訊。
5. 如申請專利範圍第 1 項所述之文件控管裝置，其中該文件為電子文件。
6. 如申請專利範圍第 5 項所述之文件控管裝置，其中該文件為電子郵件(含本文及/或附件)、傳真機傳送的電子檔、掃描裝置讀入的電子檔、及/或電腦所產生的各式電子檔案。
7. 如申請專利範圍第 1 項所述之文件控管裝置，其中該特徵標記識別器進一步具有新詞學習及用字/用語/詞性/意境的統計分析功能。
8. 如申請專利範圍第 1 項所述之文件控管裝置，其中該特徵標記識別器進一步具有資料探勘功能。
9. 如申請專利範圍第 1 項所述之文件控管裝置，其中該檔

案接收裝置為 Http、FTP、IMAP、及/或 SMTP。

10. 如申請專利範圍第 9 項所述之文件控管裝置，其中該檔案接收裝置為 FTP、IMAP、及/或 SMTP。

11. 一種具有文件控管功能的多功事務裝置，其包含：

一事務部件；

一檔案接收裝置，用以由該事務部件讀取文件；

一光學識別器，用以對該檔案接收裝置讀取的文件中的非文字內容進行光學識別；

一特徵標記識別器，用以建立該輸入文件的特徵標記；
及

一資料庫，用以儲存該讀入文件，及/或依需要將資料庫的文件輸出、經由該檔案接收裝置、或該事務部件輸出；

其特徵在：

該光學識別器可自動對該輸入文件的非文字部分進行光學辨識，得到光學辨識結果；

該特徵標記識別器，用以依該文件的特徵內容，建立該文件的特徵標記，其中該文件的特徵內容包含該文件的文字內容及/或該光學辨識的結果；

其儲存文件時，係依該檔案接收裝置讀入的來源辨識資訊及/或該文件的特徵標記進行分類，做為儲存該輸入資料的依據；及

其於儲存該文件時，依該特徵標記建立索引，用以做為裝置欲輸出文件時，搜尋該筆文件的依據。

12.如申請專利範圍第 11 項所述之多功事務裝置，其中該光學識別器為光學字符識別器。

13. 如申請專利範圍第 11 項所述之多功事務裝置，其中該裝置儲存文件時，係依該檔案接收裝置讀入的來源辨識資訊進行分類，做為儲存該輸入資料的依據。

14.如申請專利範圍第 11 項所述之多功事務裝置，其中該來源辨識資訊為檔頭資訊。

15. 如申請專利範圍第 11 項所述之多功事務裝置，其中該文件為電子文件。

16. 如申請專利範圍第 15 項所述之多功事務裝置，其中該文件為傳真機傳送的電子檔、掃描裝置讀入的電子檔、及/或電腦所產生的各式電子檔案。

17. 如申請專利範圍第 11 項所述之多功事務裝置，其中該特徵標記識別器進一步具有新詞學習及用字/用語/詞性/意境的統計分析功能。

18. 如申請專利範圍第 11 項所述之多功事務裝置，其中該特徵標記識別器進一步具有資料探勘功能。

19. 如申請專利範圍第 11 項所述之多功事務裝置，其中該檔案接收裝置為 Http、FTP、IMAP、及/或 SMTP。

20. 如申請專利範圍第 19 項所述之多功事務裝置，其中該檔案接收裝置為 FTP、IMAP、及/或 SMTP。

21. 一種文件控管方法，其包括：

一文件接收步驟，用以接收待儲存的電子文件；

一文件分解步驟，用以分解該電子文件的來源辨識資訊；

一分類步驟，用以依該來源辨識資訊進行分類；及

一檔案儲存步驟，用以依該分類儲存該電子文件；

其特徵在其進一步包括：

一特徵標記辨識步驟，用以該電子文件的內容辨識特徵標記；及

一索引建立步驟，用以依該特徵標記建立索引，作為裝置欲輸出該電子文件時，搜尋該筆該電子文件的依據。

22 如申請專利範圍第 21 項所述之文件控管方法，其中該特徵標記辨識步驟之前，進一步含一光學辨識步驟，用以辨識該電子文件中的非文字內容，並以辨識結果做為

該特徵標記辨識步驟的特徵標記辨識內容。

23. 如申請專利範圍第 22 項所述之文件控管方法，其中該光學辨識步驟，係以光學字符辨識器進行光學辨識。
24. 如申請專利範圍第 21 項所述之文件控管方法，其中該特徵標記辨識步驟之前，進一步含一光學辨識步驟，用以辨識該電子文件中的非文字內容，並將辨識結果和該電子文件的文字內容合併，做為該特徵標記辨識步驟的特徵標記辨識內容。
25. 如申請專利範圍第 24 項所述之文件控管方法，其中該光學辨識步驟，係以光學字符辨識器進行光學辨識。
26. 如申請專利範圍第 21 項所述之文件控管方法，其中該來源辨識資訊為檔頭資訊。
27. 如申請專利範圍第 21 項所述之文件控管方法，其中該特徵標記識別器進一步具有新詞學習及用字/用語/詞性/意境的統計分析功能。
28. 如申請專利範圍第 21 項所述之文件控管方法，其中該特徵標記識別器進一步具有資料探勘功能。

29. 如申請專利範圍第 21 項所述之文件控管方法，其中該檔案接收裝置為 Http、FTP、IMAP、及/或 SMTP。

30. 如申請專利範圍第 29 項所述之文件控管方法，其中該檔案接收裝置為 FTP、IMAP、及/或 SMTP。

十一、圖式：

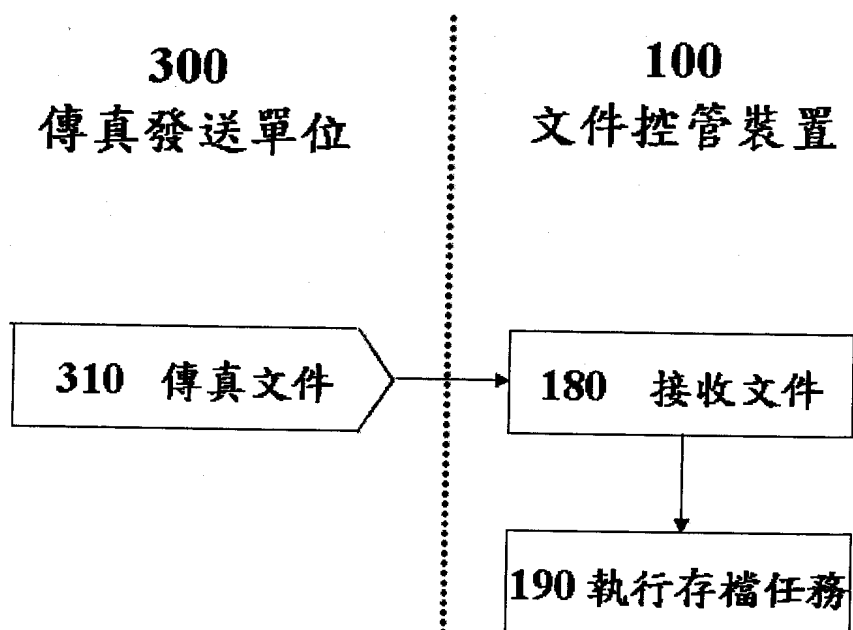


圖 1a

200 MFP 100 文件控管裝置

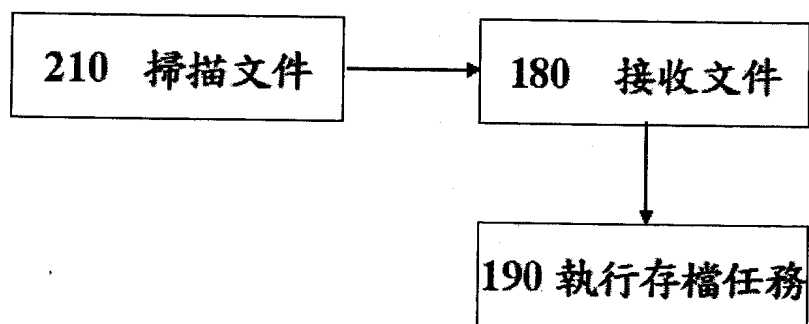


圖 1b

205 電腦

100 文件控管裝置

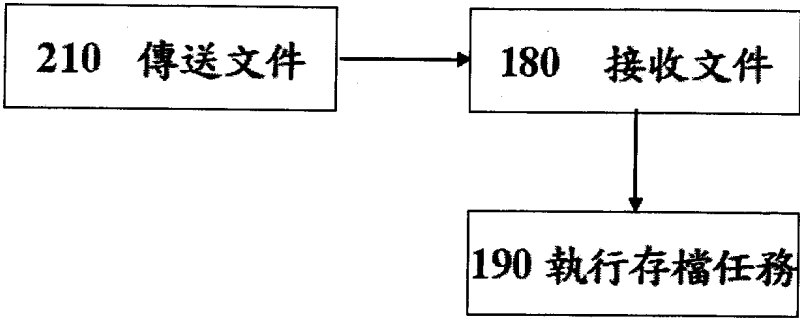


圖 1c

205 電腦

100文件控管裝置

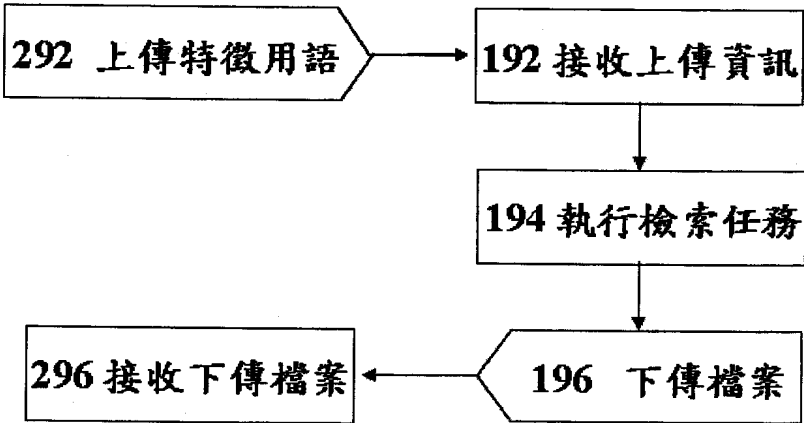


圖 2

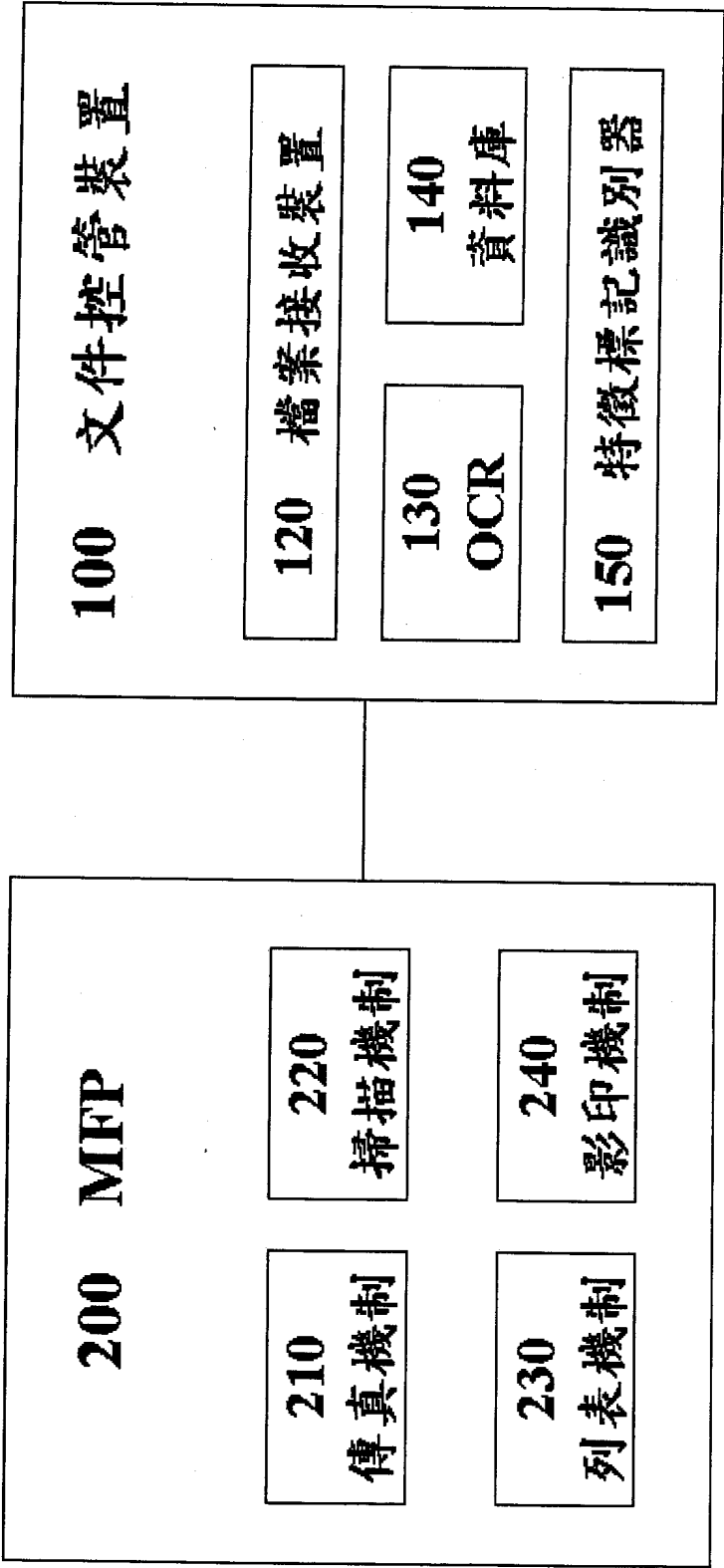


圖 3a

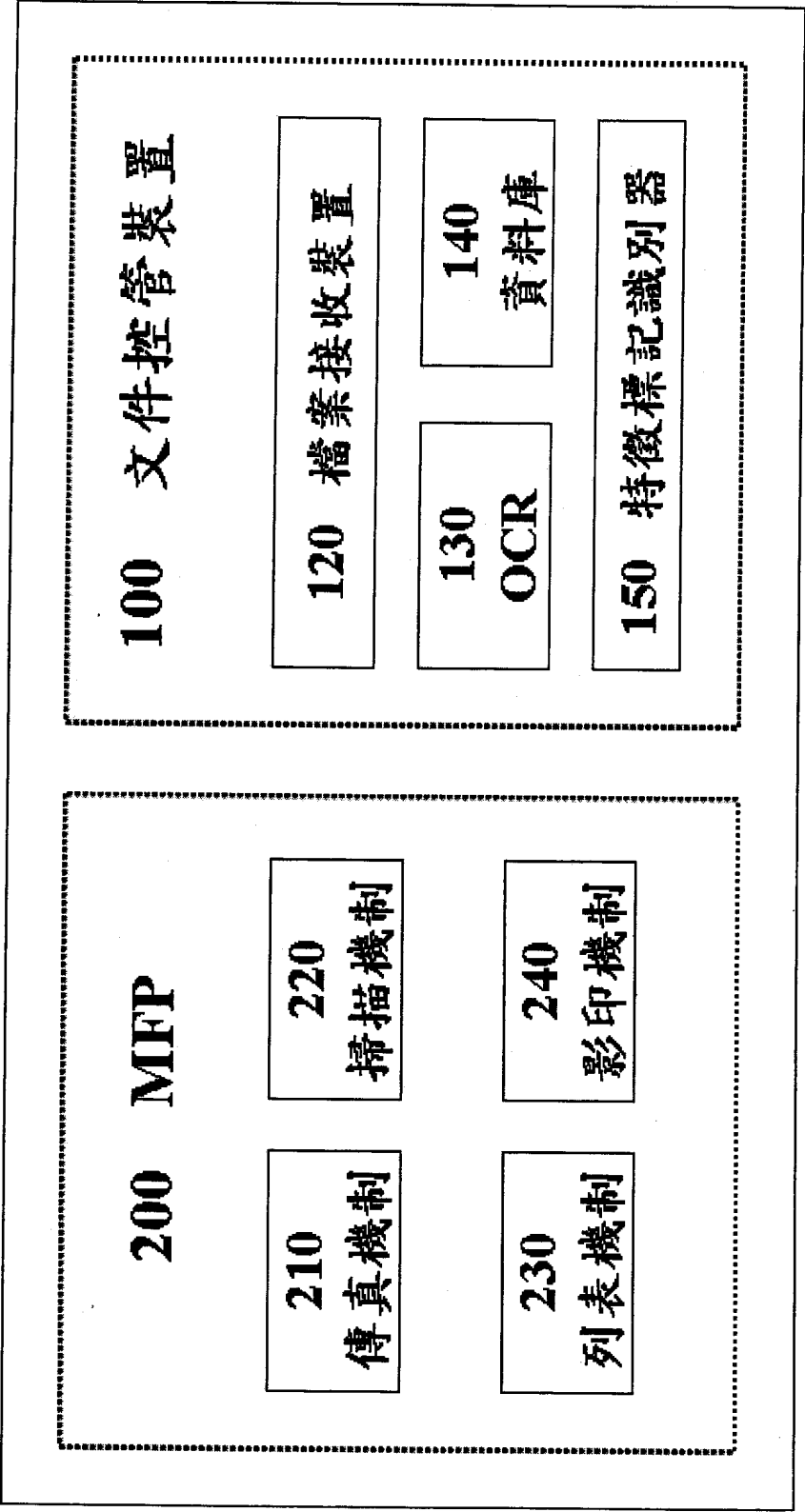


圖 3b

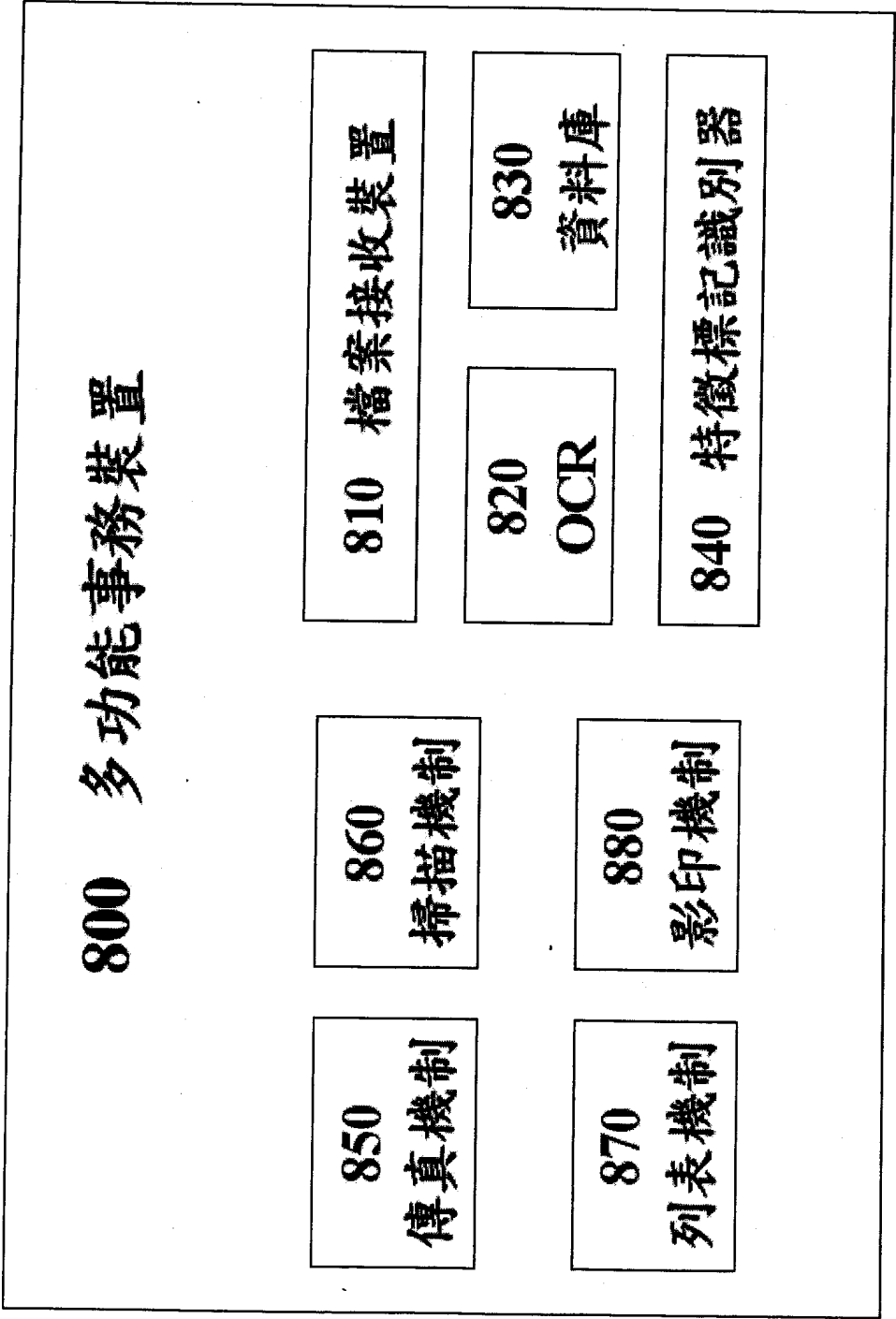


圖 3c

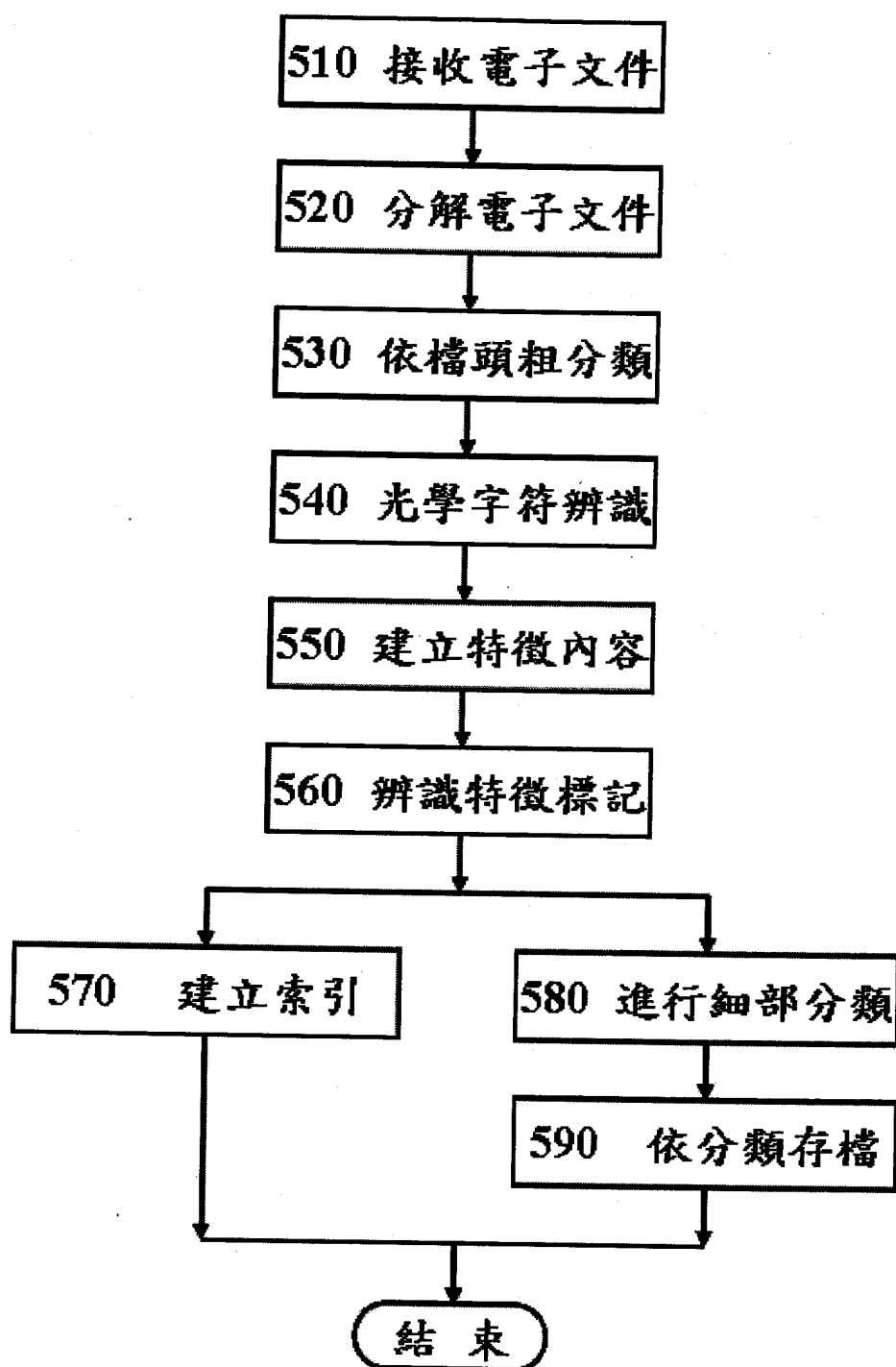


圖 4

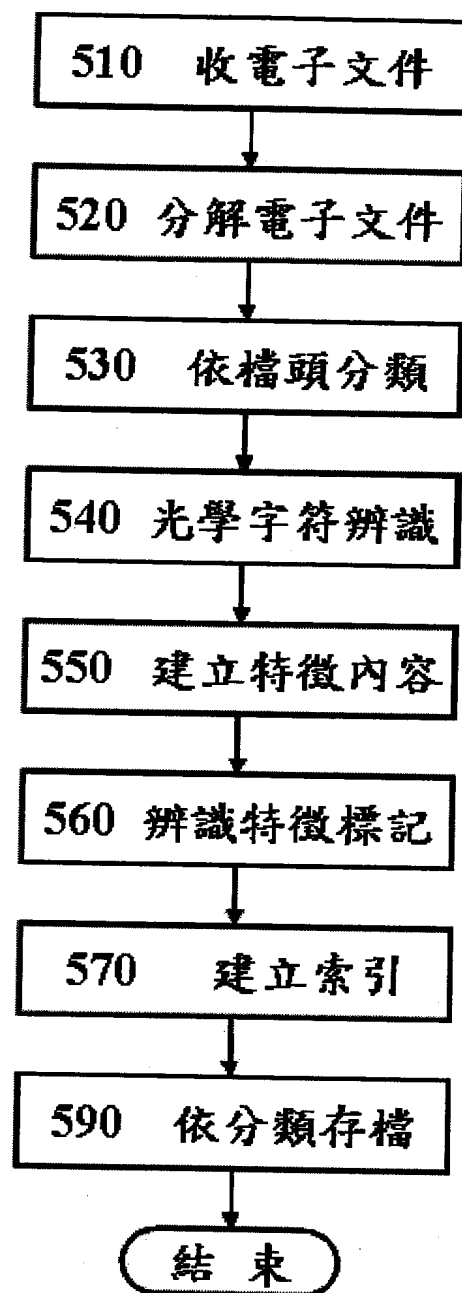


圖 5

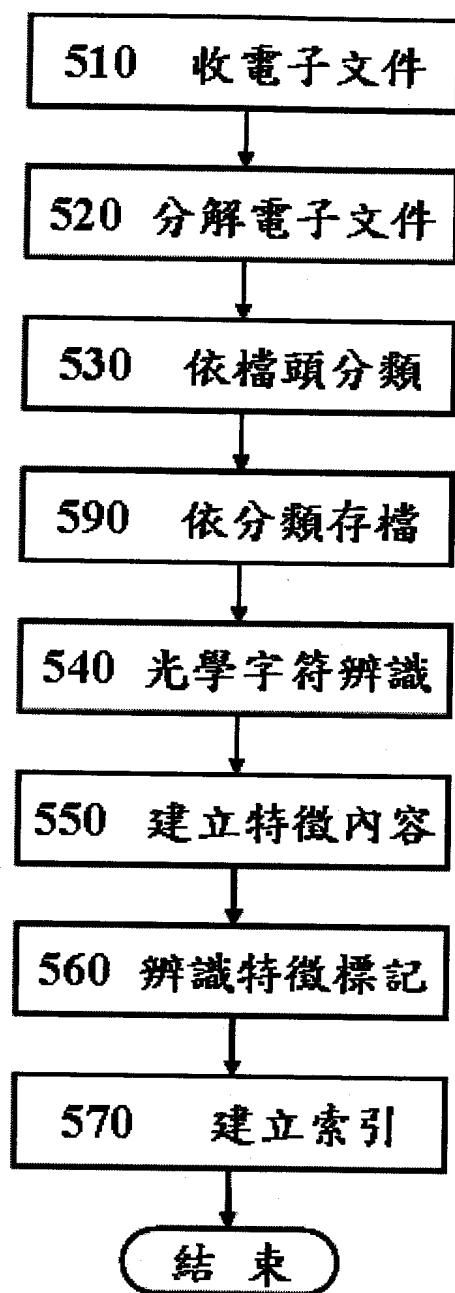


圖 6

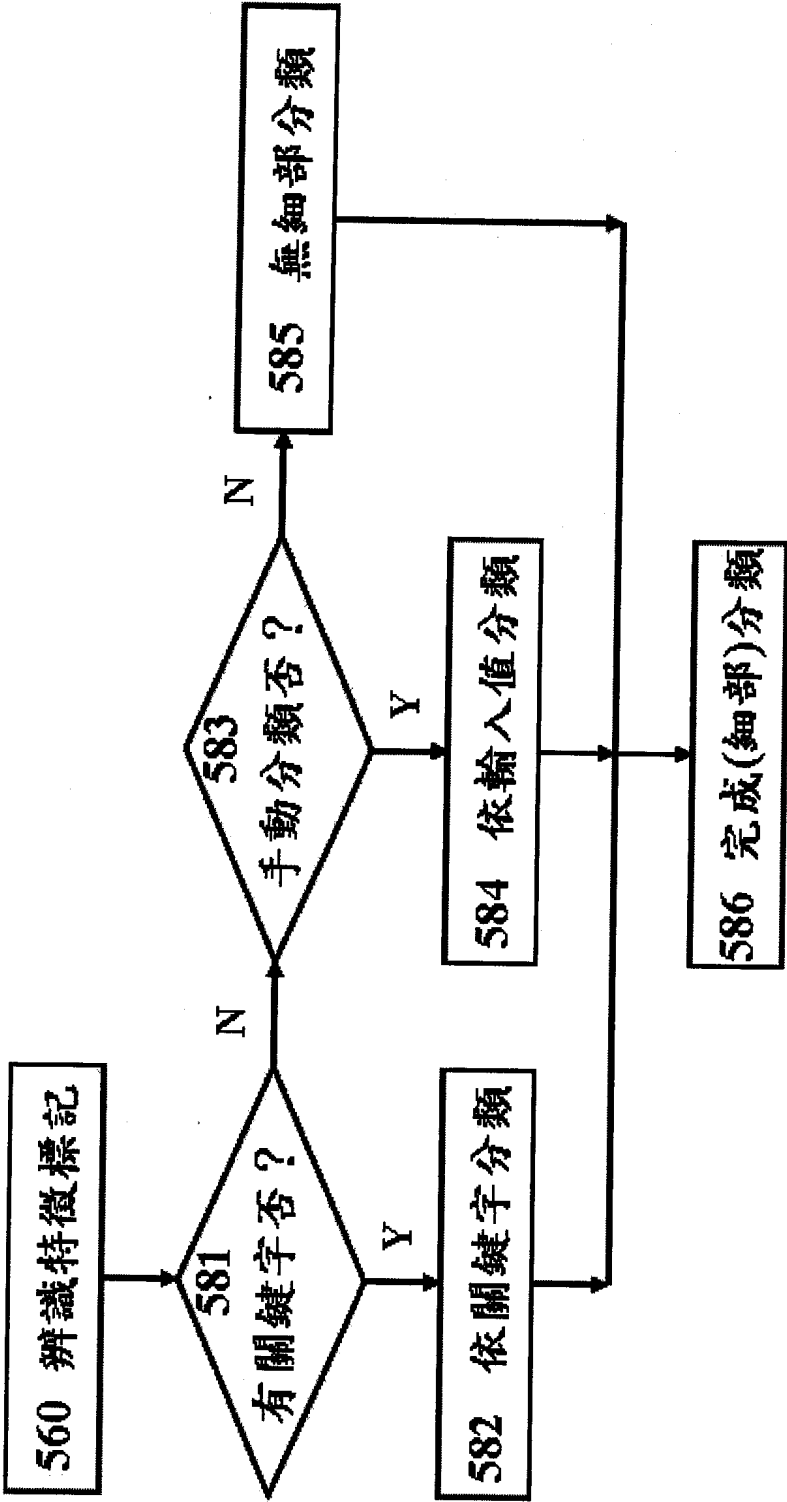


圖 7

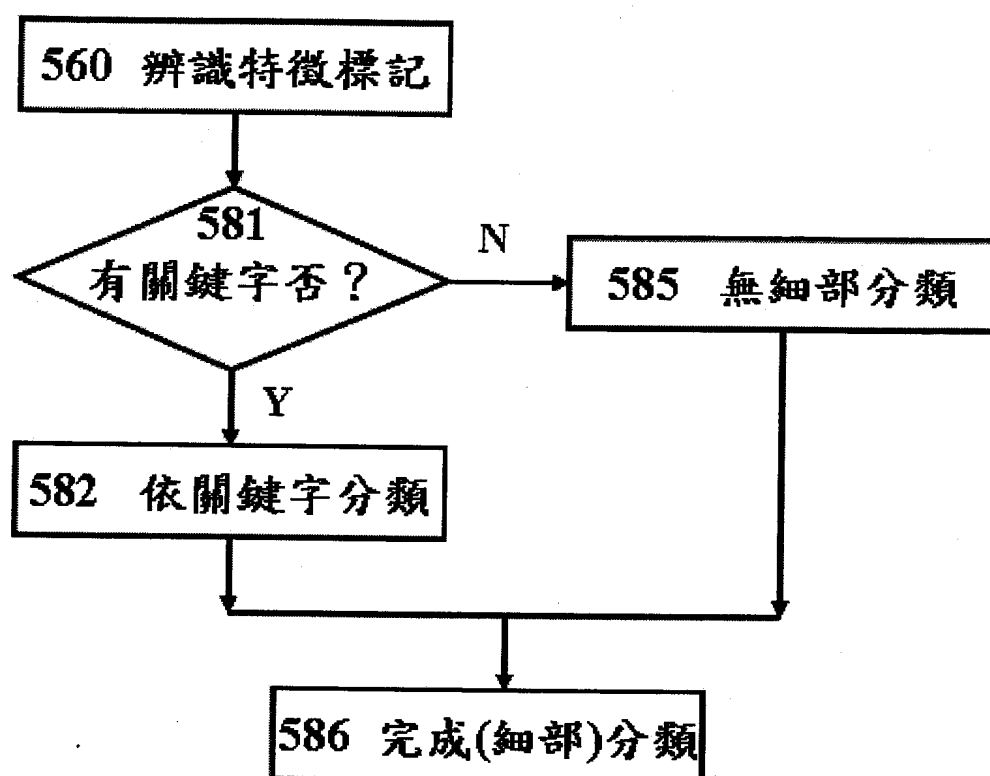


圖 8

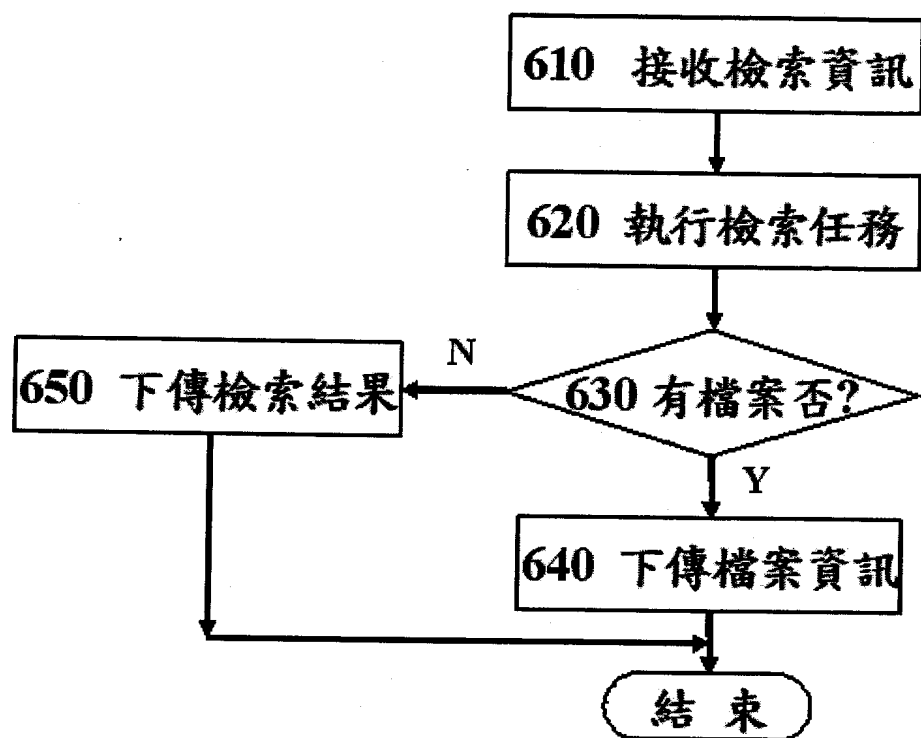


圖 9

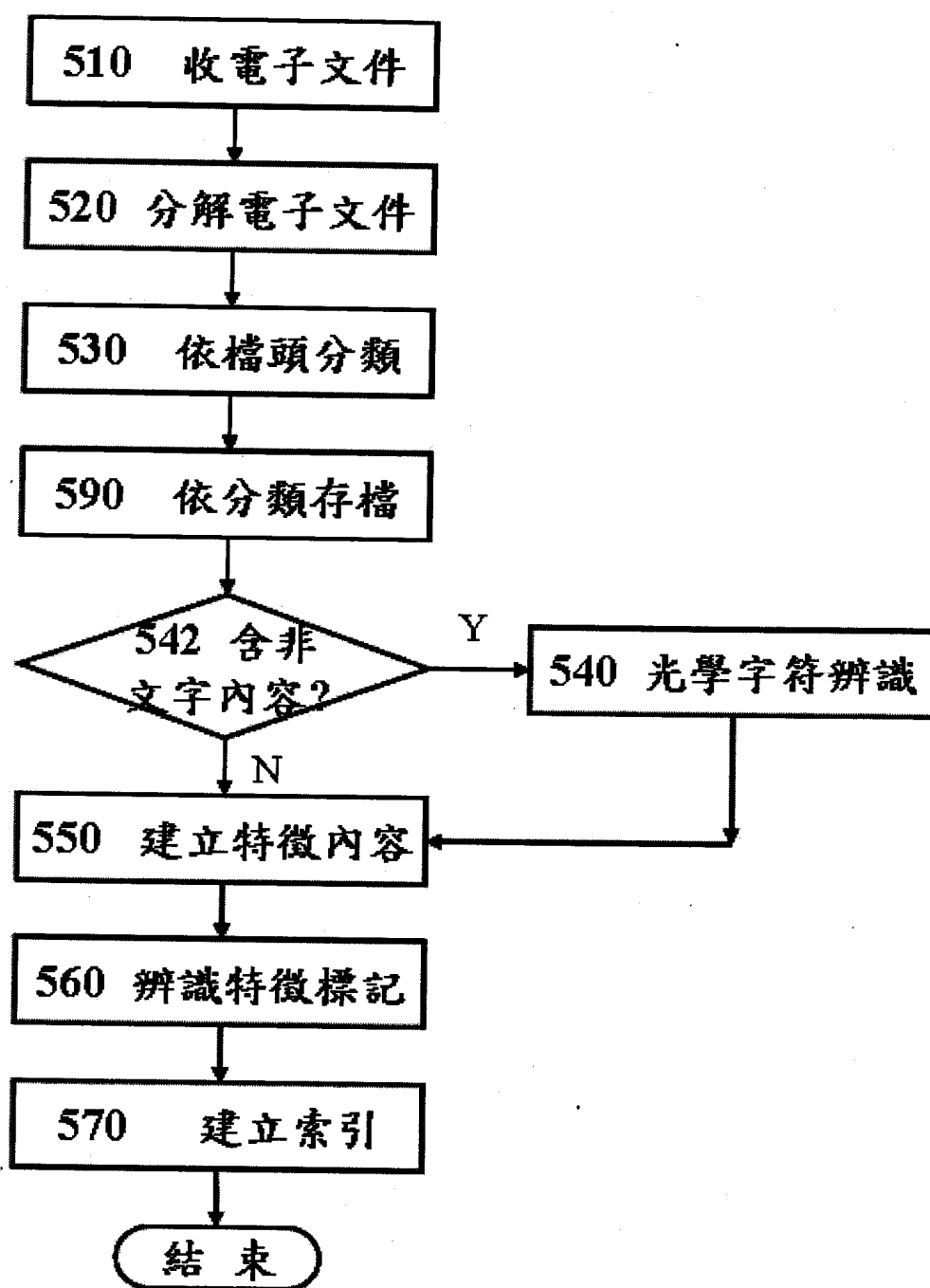


圖 10

七、指定代表圖：

(一)本案指定代表圖為：圖 3c。

(二)本代表圖之元件符號簡單說明：

800..多功事務裝置、810..檔案接收裝置、820..光學字
符識別器(OCR)、830..資料庫、840..特徵標記識別器、
850..傳真機制、860..掃描機制、870..印表機制、880..
影印機制

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

[無]

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號：97129953

※申請日期：97.8.6

※IPC 分類：G06K 9/14 (2006.01)

G06F 17/40 (2006.01)

G06F 17/30 (2006.01)

G06F 17/00 (2006.01)

一、發明名稱：(中文/英文)

具有辨識分類搜尋儲存功能的文件控管裝置及方法

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

眾來科技股份有限公司

代表人：(中文/英文) 林益邦

住居所或營業所地址：(中文/英文)

台北市濟南路二段 29-1 號

國 籍：(中文/英文) 中華民國/ROC

三、發明人：(共 4 人)

姓 名：(中文/英文)

1. 劉立恩

2. 林益邦

3. 陳衍彰

4. 林聖富

國 籍：(中文/英文) 1.-4. 中華民國/ROC