



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I525457 B

(45)公告日：中華民國 105 (2016) 年 03 月 11 日

(21)申請案號：099140209

(22)申請日：中華民國 99 (2010) 年 11 月 22 日

(51)Int. Cl. : G06F17/30 (2006.01)

G06Q30/00 (2012.01)

(30)優先權：2010/09/21 中國大陸

201010292384.0

(71)申請人：阿里巴巴集團控股有限公司(開曼群島) ALIBABA GROUP HOLDING LIMITED
(KY)

香港

(72)發明人：顧海傑(CN)；蘇寧軍(CN)；代其鋒(CN)；馬海平(CN)；張金銀(CN)；陳恩紅(CN)

(74)代理人：林志剛

(56)參考文獻：

TW 201030535A1

US 2001/0044758A1

US 2002/0026386A1

US 2004/0143600A1

US 2009/0119172A1

審查人員：郭彥鋒

申請專利範圍項數：13 項 圖式數：4 共 39 頁

(54)名稱

資訊處理的方法及設備

(57)摘要

本申請公開了一種資訊處理的方法及設備，主要內容包括：在葉子類目的基礎上，根據葉子類目下各標準產品資訊中的至少一個產品屬性參數對該標準產品資訊進行劃分，得到該葉子類目下的多個集合，由於劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同，因此，最終得到的是產品資訊細粒度劃分的集合，買家用戶以細粒度的產品資訊集合為基礎搜索、查詢產品資訊時，可以有效減少搜索、查詢的時間、提高搜索、查詢的準確性以及提高向買家用戶推薦產品資訊的準確度，且細粒度的產品資訊集合也能夠提高對產品資訊進行操作的可用性，減少對產品資訊進行操作時的運算量。

指定代表圖：

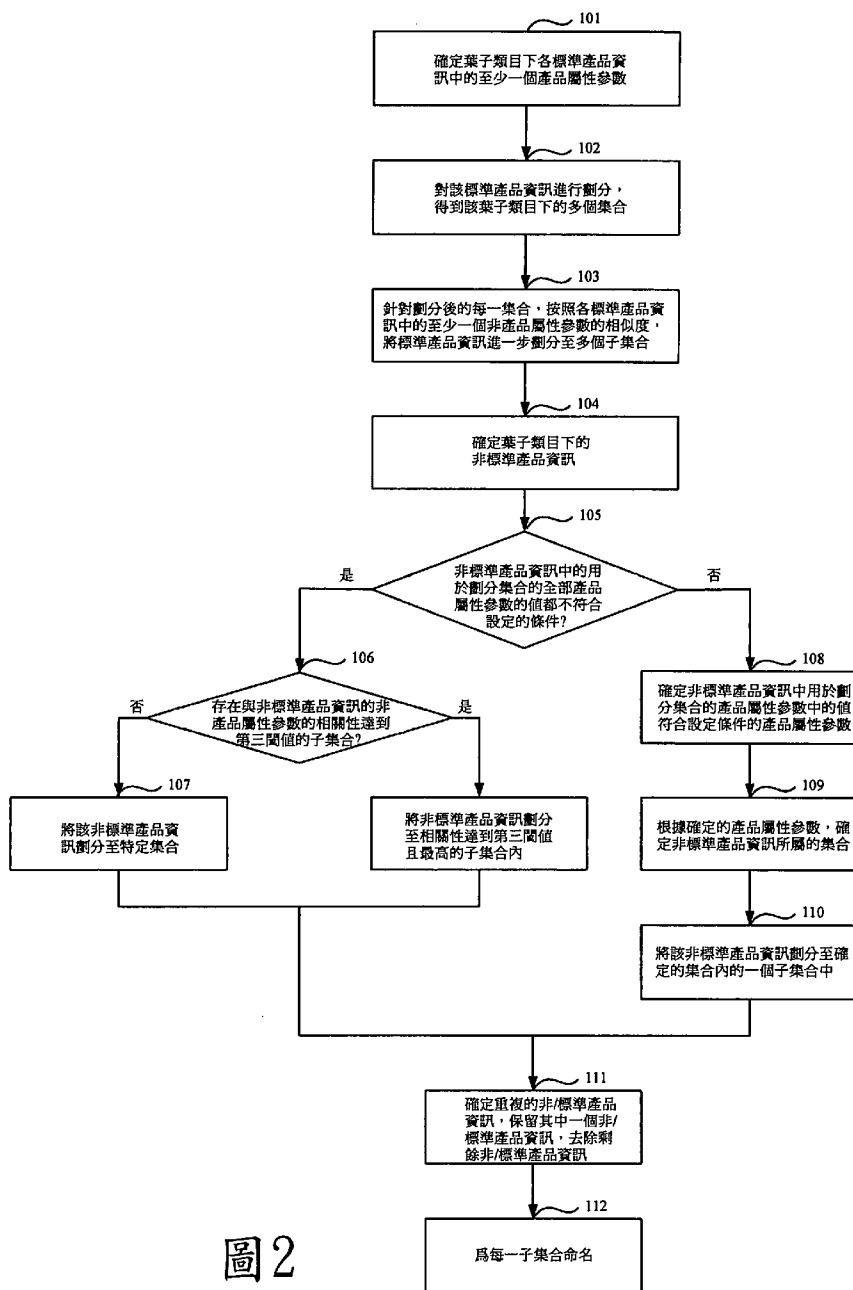


圖2

862846

發明專利說明書

(本申請書格式、順序，請勿任意更動，※記號部分請勿填寫)

※申請案號：099140209

※申請日：099 年 11 月 22 日

※IPC 分類：G06F 11/30 (2006.01)

一、發明名稱：(中文／英文)

G06Q 30/00 (2012.01)

資訊處理的方法及設備

二、中文發明摘要：

本申請公開了一種資訊處理的方法及設備，主要內容包括：在葉子類目的基礎上，根據葉子類目下各標準產品資訊中的至少一個產品屬性參數對該標準產品資訊進行劃分，得到該葉子類目下的多個集合，由於劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同，因此，最終得到的是產品資訊細粒度劃分的集合，買家用戶以細粒度的產品資訊集合為基礎搜索、查詢產品資訊時，可以有效減少搜索、查詢的時間、提高搜索、查詢的準確性以及提高向買家用戶推薦產品資訊的準確度，且細粒度的產品資訊集合也能夠提高對產品資訊進行操作的可用性，減少對產品資訊進行操作時的運算量。

三、英文發明摘要：

四、指定代表圖：

(一) 本案指定代表圖為：第(2)圖。

(二) 本代表圖之元件符號簡單說明：無

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：無

六、發明說明：

【發明所屬之技術領域】

本申請涉及電腦技術領域，尤其涉及一種資訊處理的方法及設備。

【先前技術】

隨著電腦技術以及通信技術的不斷發展，越來越多的用戶在購物網站上搜索、查詢、購買商品。用戶在搜索、查詢、購買商品之前，會瀏覽賣家用戶在網站上發佈的產品資訊，該賣家用戶可以是企業實體、廠商或是個人經營者。

網站伺服器接收到的賣家用戶上報的產品資訊是海量資訊，為了對接收到的產品資訊所表示的產品進行分類以及有效地引導買家用戶獲得其想要的產品資訊，網站伺服器通常採用多級類目的方式來劃分產品資訊。多級類目體系一般有以下兩個特徵：

特徵 1：多級類目體系的架構相對穩定。

架構相對穩定的多級類目體系一方面有助於賣家用戶積累習慣，在向網站伺服器上報產品資訊時，按照多級類目體系的要求上報標準格式和內容的產品資訊。另一方面有助於買家用戶積累習慣，根據累計的經驗在多級類目體系下快速搜索、查找想要獲得的產品資訊。

特徵 2：多級類目體系一般由網站伺服器的運營人員人工運營。

人工運營的方式可以將本領域的共有知識固定下來形成標準，有助於多級類目體系在各種網站內的推廣使用。

爲了保持多級類目體系的上述兩點特徵，在通過多級類目的方式劃分海量產品資訊時，只能按照相對較粗的粒度劃分產品資訊，這是因爲：由於產品資訊的多種多樣，如果將產品資訊劃分爲較細的粒度，則多級類目的最底層葉子類目需要隨著產品資訊的改變而變化，不利於多級類目體系的穩定；且如果細粒度地劃分產品資訊，生成的多級類目的架構勢必會非常龐大，增加了人工運營網站伺服器的難度。

例如：若某一葉子類目下是連衣裙的產品資訊，針對其中的某一產品資訊，在該產品資訊中的產品材質由絲質修改爲棉質時，該產品資訊仍是該連衣裙葉子類目下的產品資訊，葉子類目不發生變化。若多級類目體系劃分的粒度更細，如某一葉子類目下是絲質連衣裙的產品資訊，則當某一產品資訊中的產品材質由絲質修改爲棉質時，需要將該產品資訊由絲質連衣裙的葉子類目改爲棉質連衣裙的葉子類目，此時，葉子類目隨著產品資訊的改變而變化。同時，由於多級類目體系是樹狀的節點架構，因此，每增加一層子類目，多級類目體系中將增加大量的類目，使得多級類目的架構非常龐大。

由於多級類目體系下的產品資訊劃分粒度較粗，因此，即使是多級類目體系中的最底層的葉子類目中包含的產品資訊也依舊是大量的。在此情況下，買家用戶通過多級

類目體系搜索、查詢產品資訊時的查詢時間較長，且查詢的準確度較低，且網站伺服器向買家用戶推薦產品資訊時，也只能以葉子類目為單位，向買家用戶推薦葉子類目下的所有產品資訊，使得推薦的產品資訊差異很大，準確度不能滿足買家用戶的實際需求。除此以外，由於葉子類目中包含的產品信息量很大，屬於同一葉子類目中的產品資訊差異也很大，因此，針對葉子類目下的產品資訊的操作實現難度也較大。例如：在通過價格參數來自動抓取不安全的產品資訊的操作中，一般認為極端價格很可能涉嫌假冒產品，假設 A 品牌的產品價格低於 100 元時表示該產品為假冒產品，而 B 品牌的同類型產品價格低於 20 元時表示該產品為假冒產品，如果某一產品的價格為 50 元，則通過價格參數的方式很難直接定位出價格為 50 元的產品是否為假冒產品，還必須結合該產品的其他資訊來判斷，而葉子類目下的產品資訊眾多，使得運算量非常大。

【發明內容】

本申請實施例提供一種資訊處理的方法及設備，用以解決現有技術中存在的多級類目體系下產品資訊劃分粒度較大的問題。

一種資訊處理的方法，該方法包括：

確定葉子類目下各標準產品資訊中的至少一個產品屬性參數；

利用確定的該至少一個產品屬性參數對該標準產品資

訊進行劃分，得到該葉子類目下的多個集合；

其中，劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同。

一種資訊處理的設備，該設備包括：

標準參數確定模組，用於確定葉子類目下各標準產品資訊中的至少一個產品屬性參數；

第一劃分模組，用於利用確定的該至少一個產品屬性參數對該標準產品資訊進行劃分，得到該葉子類目下的多個集合，其中，劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同。

本申請有益效果如下：

本申請實施例在葉子類目的基礎上，根據葉子類目下各標準產品資訊中的至少一個產品屬性參數對該標準產品資訊進行劃分，得到該葉子類目下的多個集合，由於劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同，因此，最終得到的是產品資訊細粒度劃分的集合，買家用戶以細粒度的產品資訊集合為基礎搜索、查詢產品資訊時，可以有效減少搜索、查詢的時間、提高搜索、查詢的準確性以及提高向買家用戶推薦產品資訊的準確度，且細粒度的產品資訊集合也能夠提高對產品資訊進行操作的可用性，減少對產品資訊進行操作時的運算量。

【實施方式】

爲了實現本申請目的，本申請在多類目體系的葉子類目基礎上，對葉子類目下的產品資訊按照其產品屬性參數作進一步劃分，在每一個葉子類目下劃分出多個細粒度的產品資訊集合，使得在買家用戶搜索、查詢產品資訊時，以細粒度的產品資訊集合爲基礎，可以有效減少搜索、查詢的時間、提高搜索、查詢的準確性以及提高網站伺服器向買家用戶推薦產品資訊的準確度，且細粒度的產品資訊集合也能夠提高對產品資訊進行操作的可用性，減少對產品資訊進行操作時的運算量；並且，本申請方案是在多類目體系的葉子類目基礎上執行的，對多級類目體系的實質性內容沒有改變，多級類目體系本身仍然具有上述的兩點特徵。

本申請各實施例中涉及的資訊處理設備可以應用在多級類目體系下，其架構如圖 1 所示，在多級類目體系的最底層葉子類目下，資訊處理設備對葉子類目中的產品資訊進行處理，得到葉子類目下一層次的對產品資訊做細粒度劃分的集合和子集合。多級類目體系下的各葉子類目是相對獨立的類目，本申請方案是要在每一葉子類目下繼續構建細粒度的類目結構，因此，本實施例方案中可以採用分散式演算法，將每一個葉子類目作爲一個計算節點，通過電腦集群對多個葉子類目進行分散式操作，以加快本實施例中的資訊處理方案。在圖 1 所示的結構下，獨立於多級類目體系下的一個資訊處理設備可以對多級類目體系下的多個葉子類目中的產品資訊進行處理。

若將本申請方案中的資訊處理方案處理後的集合和子集合中的標準產品資訊應用於向買家用戶推薦產品資訊或是買家用戶的產品資訊搜索的場景，則用於向買家用戶推送產品資訊的伺服器根據買家用戶的需求（如買家用戶輸入的關鍵字或是買家用戶在之前一段時間內的購買習慣），將集合或子集合中的產品資訊推送給買家用戶，或是搜索引擎從集合和子集合中進行關鍵字搜索，並將搜索得到的產品資訊發送給買家用戶。

根據葉子類目下的產品資訊的產品屬性參數值是否符合設定的條件，可以將其劃分為標準產品資訊和非標準產品資訊，標準產品資訊的產品屬性參數值符合設定的條件，而非標準產品資訊的某些或全部產品屬性參數值不符合設定的條件。該設定的條件可以是產品屬性參數值的取值是實際可用的取值。例如：某一葉子類目下包含的是針對連衣裙的產品資訊，產品資訊 1 中的價格參數值是 100～150 元區間，而產品資訊 2 中的價格參數值是 ABC，並不是表示價格的正數數值，說明賣家用戶在填寫產品資訊 2 中的價格參數值時可能出現失誤，則將產品資訊 2 看作是非標準產品資訊。

不論是標準產品資訊還是非標準產品資訊都包括產品屬性參數和非產品屬性參數，產品屬性參數表示該產品資訊對應產品本身的固定屬性，包括但不限於：產品的價格、產品的材質、產品的品牌、產品的型號、產品的重量等；非產品屬性參數表示賣家用戶或是網站伺服器為產品定

義的屬性，包括但不限於：產品資訊所屬的葉子類目、發佈產品的標題、賣家用戶 ID、產品的用途等。

下面結合說明書附圖對本申請實施例進行詳細描述。

實施例一：

如圖 2 所示，為本申請實施例一中資訊處理方法示意圖，該方法包括以下步驟：

步驟 101：確定葉子類目下各標準產品資訊中的至少一個產品屬性參數。

假設本步驟中的葉子類目包含的是針對連衣裙的產品資訊，包含的標準產品資訊有 1000 個，其中 3 個標準產品資訊（標準產品資訊 1、標準產品資訊 2 和標準產品資訊 3）的內容如表 1 所示：每個標準產品資訊中有 4 個產品屬性參數，分別是連衣裙的價格、連衣裙的材質、連衣裙的品牌、連衣裙的型號。

標準產品資訊	價格	材質	品牌	型號
標準產品資訊 1	價格區間 1	材質 1	A	型號 1
標準產品資訊 2	價格區間 1	材質 1	B	型號 2
標準產品資訊 3	價格區間 2	材質 2	C	型號 3

表 1

若將表 1 中的 4 個產品屬性參數都作為細粒度產品資訊的劃分條件，對葉子類目下的標準產品資訊進行劃分，則會由於劃分條件過於複雜導致劃分後得到的集合數量過

多。例如：若葉子類目下的所有 1000 個標準產品資訊中共包含 4 種連衣裙的價格區間、3 種連衣裙的材質、50 種連衣裙的品牌、5 種連衣裙的型號，則最多將會劃分得到 $4 \times 3 \times 50 \times 5 = 3000$ 個集合。而在實際的處理過程中，產品屬性參數的種類可能會更多，每種產品屬性參數值的可選項也更多，因此，將全部產品屬性參數都作為劃分條件的話，運算量會比較大。對此，本步驟提出一種優化方案，從全部產品屬性參數中選擇部分產品屬性參數來作為本步驟 101 中確定的產品屬性參數用於作細粒度的產品資訊劃分。

具體的選擇產品屬性參數的操作需要滿足以下兩方面要求：

一方面，選擇的產品屬性參數應該是對外最能反映該產品資訊的參數，即買家用戶常用於搜索、查詢的參數，以便於根據選擇的產品屬性參數進行劃分後得到的集合能夠提高用戶搜索、查詢的準確性；另一方面，選擇的產品屬性參數的值最好是離散型的，以減少在後續劃分集合時由於標準產品資訊中的產品屬性參數的取值可能性過多導致運算量大的問題。

為了滿足以上兩方面要求，本步驟中選擇用戶集合劃分的產品屬性參數的條件是：產品屬性參數作為搜索條件，在之前的一段設定時長內，在葉子類目下進行產品資訊搜索的次數需要達到第一閾值，且該產品屬性參數的值為離散型。

按照上述條件選擇出至少一個產品屬性參數後用於後續的集合劃分操作。仍以表 1 所示的產品資訊中的產品屬性參數為例，若葉子類目下的所有 1000 個標準產品資訊中共包含 4 種連衣裙的價格區間、3 種連衣裙的材質、50 種連衣裙的品牌、5 種連衣裙的型號，則可以將其中連衣裙的價格的值的 4 個區間看作離散型；連衣裙的材質的取值為 3 種不同的材質，可以看作是離散型；連衣裙的品牌的取值為 50 種不同的品牌，可以看作是離散型；連衣裙的型號的取值為 5 種不同的型號，可以看作是離散型。若買家用戶最常用價格和材質為搜索條件進行產品資訊的搜索、查詢，則將產品的價格、產品的材質作為本步驟中確定的用於集合劃分的產品屬性參數。

步驟 102：對該標準產品資訊進行劃分，得到該葉子類目下的多個集合，其中，劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同。

在本步驟中，若用於劃分集合的產品屬性參數是產品的價格和產品的材質，其中，產品的價格有 4 種價格區間，產品的材質有 3 種不同的材質，則劃分後應該得到以下 12 個集合：

集合 1：價格區間 1+材質 1；集合 2：價格區間 1+材質 2；集合 3：價格區間 1+材質 3；集合 4：價格區間 2+材質 1；集合 5：價格區間 2+材質 2；集合 6：價格區間 2+材質 3；集合 7：價格區間 3+材質 1；集合 8：價格區間 3+材質 2；集合 9：價格區間 3+材質 3；集合 10：價格

區間 4+材質 1；集合 11：價格區間 4+材質 2；集合 12：價格區間 4+材質 3。

劃分在同一集合中的任意兩個標準產品資訊中的價格區間和材質都相同，根據表 1 所示的各標準產品資訊的產品屬性參數描述，標準產品資訊 1 和標準產品資訊 2 應劃分在集合 1，標準產品資訊 3 應劃分在集合 5，任意兩個集合中滿足類與類不相似的原則。

由於在步驟 101 中選擇用於細粒度產品資訊劃分的產品屬性參數時，充分考慮到了買家用戶的搜索、查詢習慣以及產品屬性參數的值為離散型，同時忽略未選擇的產品屬性參數，使得本步驟的劃分結果不僅能夠正確體現買家用戶的使用習慣，還減少了劃分時的運算量。

通過以上步驟 101 和步驟 102 的方案，對多級類目體系的葉子類目中的產品資訊作了細粒度的劃分，得到了細粒度劃分產品資訊的集合，使得買家用戶通過集合內的細粒度產品進行搜索、查詢時可以有效減少等待時間、提高準確性以及提高網站伺服器向買家用戶推薦產品資訊的準確性，有助於提高對產品資訊的各項操作的可行性。例如，在將連衣裙的葉子類目劃分為上述 12 個集合後，可以根據買家用戶的搜索條件在相應的集合內進行搜索，由於集合內的產品資訊數量遠遠小於葉子類目下的產品資訊數量，因此，可以大大減少買家用戶搜索等待時間，且保證搜索、查詢的準確性；同時，在向買家用戶推薦產品資訊時，根據向買家用戶推薦特定價位以及材質的產品資訊，

使得推薦的產品資訊接近買家用戶實際需求；另外，仍以通過價格參數來自動抓取不安全的產品資訊的操作為例，通過上述方式劃分細粒度的產品資訊集合後，可以根據待測的價格快速定位出相應的集合，進而在定位出的集合中查詢是否存在假冒產品，相比於在葉子類目下的操作，可以有效地減少操作的運算量，提高操作執行的有效性。

在上述步驟 101 和步驟 102 的優選方案中，根據最能反映產品資訊以及取值為離散型的產品屬性參數作為劃分集合的產品屬性參數，可以進一步根據劃分後得到的集合來提高用戶搜索、查詢的準確性，以及減少在劃分集合時由於標準產品資訊中的產品屬性參數的值的的可能性過多導致運算量大的問題。

步驟 103：針對劃分後的每一集合，按照集合中各標準產品資訊中的至少一個非產品屬性參數的相似度，將集合內的標準產品資訊進一步劃分至多個子集合。

本步驟是實現本申請目的的優選步驟，在葉子類目下劃分出多個集合後，再針對每一個集合作進一步的劃分，可以在步驟 102 的基礎上得到更細粒度的標準產品資訊的子集合。

本步驟的具體做法為：

首先，針對劃分後得到的每一集合，確定集合中各標準產品資訊中的至少一個非產品屬性參數。

仍以步驟 101 和步驟 102 中涉及的連衣裙產品資訊為例，假設本步驟中確定的非產品屬性參數是賣家用戶在網

站伺服器上發佈產品的標題，標準產品資訊 1 和標準產品資訊 2 劃分在集合 1 中，標準產品資訊 1 內發佈產品的標題中的關鍵字為：連衣裙、動物圖案，標準產品資訊 2 內發佈產品的標題中的關鍵字為：連衣裙、條紋圖案。

然後，確定同一集合內各標準產品資訊的至少一個非產品屬性參數的相似度。

由於以發佈產品的標題為非產品屬性參數，因此，本步驟中需要運算同一集合內各發佈產品的標題之間的相似度。如通過 K-中心點等聚類演算法運算相似度。

最後，將同一集合內各標準產品資訊做進一步劃分，得到該集合下的多個子集合，使得劃分在同一子集合內的兩兩標準產品資訊中的該至少一個非產品屬性參數之間的相似度達到第二閾值。

若通過 K-中心點等聚類演算法運算相似度時，將相似度達到第二閾值的各標準產品資訊劃分在同一子集合，不同子集合之間，非產品屬性參數的相似度較低。

本實施例也不限於通過產品資訊的其他非產品屬性參數之間的相似度來劃分子集合，如通過產品用途劃分等。

通過上述步驟 101～步驟 103 的方案，完成了針對標準產品資訊的細粒度劃分，進一步地，還可以對非標準產品資訊作細粒度的劃分。

步驟 104：確定葉子類目下的非標準產品資訊。

本步驟也是實現本申請目的的優選步驟，雖然葉子類目下的非標準產品資訊中的部分或全部產品屬性參數的值

不符合設定的條件，但如果這些值是賣家用戶上報時填寫失誤或是其他人為誤差造成的，則該非標準產品資訊也應該真實表示一個產品的相關資訊，且該非標準產品資訊也具有搜索、查詢、向用戶推薦或是用於其他操作的意義，因此，本優選步驟就是在已劃分標準產品資訊的集合和子集合後，進一步對非標準產品資訊的劃分。

步驟 105：判斷非標準產品資訊中的用於劃分集合的全部產品屬性參數的值是否都不符合設定的條件，若是，則執行步驟 106；否則執行步驟 108。

非標準產品資訊的某些或全部產品屬性參數值不符合設定的條件，在本步驟中，需要判斷在步驟 101 中確定的用於劃分集合的產品屬性參數是否是不符合設定的條件的參數。例如，在步驟 101 中確定的用戶劃分集合的產品屬性參數為產品的價格、產品的材質，則在本步驟中將確定非標準產品資訊中的產品的價格和產品的材質這兩種產品屬性參數是否都不符合設定的條件。

需要說明的是，若在步驟 101 中將產品資訊的全部產品屬性參數都用於劃分集合，則只要有產品屬性參數值是不符合設定的條件的產品資訊就定義為非標準產品資訊；若在步驟 101 中設定部分產品屬性參數用於劃分集合，則在本步驟中涉及的非標準產品資訊是指用於劃分集合的至少一個產品屬性參數值為不符合設定的條件的產品資訊。

特殊地，若存在某一非標準產品資訊，該非標準產品資訊中取值不符合設定的條件的產品屬性參數並未用於劃

分集合，如用於劃分集合的產品屬性參數為產品的價格、產品的材質，但該非標準產品資訊中取值不符合設定的條件的產品屬性參數是產品型號，則該非標準產品資訊可以作為標準產品資訊的特例，在步驟 102 中進行劃分；也可以仍舊作為非標準產品資訊，在後續步驟中劃分。

步驟 106：確定非標準產品資訊中的至少一個非產品屬性參數，並判斷是否存在與非標準產品資訊的非產品屬性參數的相關性達到第三閾值的子集合，若存在，則將非標準產品資訊劃分至相關性達到第三閾值且最高的子集合內，並跳轉至步驟 111；否則，執行步驟 107。

由於非標準產品資訊中的產品屬性參數無法用於集合劃分，因此，本步驟中利用非標準產品資訊的非產品屬性參數來判斷該非標準產品資訊應該屬於哪一子集合。

仍以用於劃分集合的產品屬性參數為產品的價格、產品的材質為例，用於劃分子集合的非產品屬性參數為發佈產品的標題，則在本步驟中，某一非標準產品資訊的產品的價格和產品的材質的值不符合設定的條件，確定該非標準產品資訊的發佈產品的標題，通過 K-中心點等聚類演算法運算該發佈產品的標題與每一子集合中的各標準產品資訊的發佈產品的標題的相似度，為減少運算量，可以與每一子集合中的一個標準產品資訊的發佈產品的標題進行相似運算，查找出相似度達到第三閾值的標準產品資訊所在的子集合，並將該非標準產品資訊劃分為相似度最高的標準產品資訊所在的子集合。

步驟 107：將該非標準產品資訊劃分至特定集合，並跳轉至步驟 111。

在本實施例中，由於存在某些非標準產品資訊的用於劃分集合的產品屬性參數的值不符合設定的條件且非產品屬性參數不與任何集合內的子集合相關，則為這一類非標準產品資訊單獨設置一個特定集合。

該特定集合與步驟 102 中劃分的集合滿足類與類之間不相似的原則。

步驟 108：確定非標準產品資訊中用於劃分集合的產品屬性參數中的值符合設定的條件的產品屬性參數。

在本步驟中，由於非標準產品資訊中部分用於劃分集合的產品屬性參數中的值符合設定的條件，因此，可以利用值符合設定的條件的這部分產品屬性參數來劃分非標準產品資訊。

例如：若在步驟 101 中確定的用於劃分集合的產品屬性參數為產品的價格、產品的材質，而某一非標準產品資訊的產品的價格的值不符合設定的條件，但產品的材質的值符合設定的條件，則可以利用產品材質這一產品屬性參數來劃分非標準產品資訊。

步驟 109：根據步驟 108 確定的產品屬性參數，確定非標準產品資訊所屬的集合。

在本步驟中，根據確定的產品屬性參數的值以及各集合中的標準產品資訊中的該產品屬性參數的值，查找出與該非標準產品資訊的產品屬性參數的值相同的集合。

確定的集合數量可能不止一個，在此情況下，可以將確定的所有集合執行後續步驟。例如：在步驟 101 中示例中產生了 12 個集合，假設本步驟中的產品屬性參數是產品的材質，取值為材質 1，則本步驟確定的該非標準產品資訊所屬的集合可能為集合 1、集合 4、集合 7 以及集合 10 中的一個。

步驟 110：根據非標準產品資訊的非產品屬性參數，將該非標準產品資訊劃分至步驟 109 確定的集合內的一個子集合中。

通過步驟 109 確定了非標準產品資訊應該屬於的集合，但還不能最終確定該非標準產品資訊應該屬於哪一集合中的哪一子集合，因此，本步驟進一步採用非產品屬性資訊對非標準產品資訊作進一步劃分，將非標準產品資訊劃分至非產品屬性參數的相似度最高的子集合中。

在步驟 106 中劃分非標準產品資訊時，與劃分至的子集合之間的相似度需要達到第三閾值，進而再選擇相似度最高的子集合，而在本步驟中只需要從步驟 109 確定的集合中選擇相似度最高的子集合即可，這是因為：在步驟 106 中，非標準產品資訊的用於劃分集合的產品屬性參數都不符合設定的條件，因此，如果只選擇與非標準產品資訊的非產品屬性參數相關性最高的子集合，而不為相關性設置最低的門限值（即第三閾值），則可能出現非標準產品資訊與任一子集合的相關性都很低，但仍舊選擇一個子集合作為非標準產品資訊歸屬的子集合，導致該非標準產

品資訊和同一子集合內的其他標準產品資訊的相關性低。而在步驟 110 中，由於在步驟 109 中確定的集合是通過非標準產品資訊的一個取值符合設定的條件的產品屬性參數來確定的，因此，在步驟 110 中用於選擇的集合與非標準產品資訊有一定的相關性，進而從中選擇的子集合與非標準產品資訊的相關性也較高。

步驟 111：確定重複的標準產品資訊，並保留其中一個標準產品資訊，去除剩餘的標準產品資訊，以及確定重複的非標準產品資訊，並保留其中一個非標準產品資訊，去除剩餘的非標準產品資訊。

若兩個標準產品資訊間的產品屬性參數和非產品屬性參數都相同，則定義這兩個標準產品資訊是重複的標準產品資訊。

若兩個非標準產品資訊間的產品屬性參數和非產品屬性參數都相同，則定義這兩個非標準產品資訊是重複的非標準產品資訊。

步驟 111 是實現本申請目的的優選步驟，由於賣家用戶在網站伺服器上可能出現重複發佈產品資訊的情況，為了減少對重複產品資訊進行處理所佔用的資源，步驟 111 對標準產品資訊和非標準產品資訊進行去重操作。

步驟 111 的去重操作可以在步驟 101 之前，或是步驟 101～步驟 110 之間的任意時刻執行，或是在步驟 110 之後執行。

步驟 112：為每一子集合命名。

在執行到步驟 110 時，實際上已經自動對產品資訊進行了劃分，得到了以集合為單位或是子集合為單位的標準產品單元（standard product unit，SPU）。本步驟作為本實施例的優選步驟，是為每一個 SPU 進行命名或者打標籤，也就是標識每一 SPU 代表的內容。

本步驟的具體做法是：

首先確定每一子集合的標準產品資訊和非標準產品資訊中的產品屬性參數和非產品屬性參數；

然後統計確定該產品屬性參數和非產品屬性參數中的至少一個高頻詞；

最後將該高頻詞用於為該子集合命名。

後續買家用戶搜索、查詢以及向買家用戶推薦產品資訊時，可以通過子集合名稱中涉及的高頻詞作為關鍵字進行搜索或推薦。

步驟 112 可以在步驟 110 之後且步驟 111 之前執行。

實施例二：

本申請實施例二提供一種與實施例一屬於同一發明構思下的資訊處理的設備，如圖 3 所示，該設備包括標準參數確定模組 11 和第一劃分模組 12，其中：標準參數確定模組 11 用於確定葉子類目下各標準產品資訊中的至少一個產品屬性參數；第一劃分模組 12 用於利用確定的該至少一個產品屬性參數對該標準產品資訊進行劃分，得到該葉子類目下的多個集合，其中，劃分在同一集合內的標準

產品資訊中的該至少一個產品屬性參數的值都相同。

如圖 4 所示，該標準參數確定模組具體包括屬性參數確定子模組 21、次數確定子模組 22 和選擇子模組 23，其中：屬性參數確定子模組 21 用於確定葉子類目下各標準產品資訊中的全部產品屬性參數；次數確定子模組 22 確定設定時長內每一產品屬性參數作為搜索條件，在該葉子類目下進行產品資訊搜索的次數；選擇子模組 23 用於從該全部產品屬性參數中選擇至少一個產品屬性參數，其中，選擇的任一產品屬性參數作為搜索條件進行產品資訊搜索的次數達到第一閾值且該產品屬性參數的值為離散型。

該設備還包括非標準參數確定模組 13、相似度確定模組 14 和第二劃分模組 15，其中：非標準參數確定模組 13 用於針對劃分後得到的每一集合，確定集合中各標準產品資訊中的至少一個非產品屬性參數；相似度確定模組 14 用於確定同一集合內各標準產品資訊的至少一個非產品屬性參數的相似度；第二劃分模組 15 用於按照確定的相似度對同一集合內各標準產品資訊做進一步劃分，得到該集合下的多個子集合，其中，劃分在同一子集合內的任意兩個標準產品資訊的該至少一個非產品屬性參數之間的相似度達到第二閾值。

該設備還包括判斷模組 16，用於判斷非標準產品資訊中的用於劃分集合的全部產品屬性參數的值是否都不符合設定的條件，若是，則觸發非標準參數確定模組 13，否則，觸發標準參數確定模組 11，該非標準產品資訊中的用於

劃分集合的至少一個產品屬性參數的值不符合設定的條件。

根據判斷模組 16 的觸發，非標準參數確定模組 13 和標準參數確定模組 11 各自的運行過程如下：

在非標準產品資訊中的用於劃分集合的全部產品屬性參數的值都不符合設定的條件時：

該非標準參數確定模組 13 還用於確定該非標準產品資訊中的至少一個非產品屬性參數；該相似度確定模組 14 用於確定非標準產品資訊的非產品屬性參數與各子集合內的標準產品資訊中的非產品屬性參數的相似度；該第二劃分模組 15 還用於根據非產品屬性參數，將該非標準產品資訊劃分至一個子集合內，其中，該非標準產品資訊的非產品屬性參數與劃分至的子集合內的標準產品資訊中的非產品屬性參數的相似度達到第三閾值。

特殊地，非標準產品資訊中的用於劃分集合的全部產品屬性參數的值不符合設定的條件，且該非標準產品資訊的非產品屬性參數與任一子集合內的標準產品資訊中的非產品屬性參數的相似度低於第三閾值時，該第一劃分模組 12 還用於在非標準產品資訊中的用於劃分集合的全部產品屬性參數的值不符合設定的條件，且該非標準產品資訊的非產品屬性參數與任一子集合內的標準產品資訊中的非產品屬性參數的相似度低於第三閾值時，將該非標準產品資訊劃分至特定集合。

在非標準產品資訊中的用於劃分集合的全部產品屬性

參數的值不全不符合設定的條件時：

該標準參數確定模組 11 還用於確定非標準產品資訊中用於劃分集合的產品屬性參數中值符合設定的條件的產品屬性參數；該第一劃分模組 12 還用於確定包含該值符合設定的條件的產品屬性參數的集合；該第二劃分模組 15 還用於根據非標準產品資訊的非產品屬性參數，將該非標準產品資訊劃分至第一劃分模組 12 確定的集合內的一個子集合中，其中，該非標準產品資訊的非產品屬性參數與劃分至的子集合內的標準產品資訊中的非產品屬性參數的相似度最高。

該設備還包括去重模組 17，用於確定重複的標準產品資訊，保留其中一個標準產品資訊，去除剩餘的標準產品資訊，以及，確定重複的非標準產品資訊，保留其中一個非標準產品資訊，去除剩餘的非標準產品資訊，該重複的標準產品資訊間的產品屬性參數和非產品屬性參數都相同，該重複的非標準產品資訊間的產品屬性參數和非產品屬性參數都相同。

該設備還包括命名模組 18，用於確定每一子集合的標準產品資訊和非標準產品資訊中的產品屬性參數和非產品屬性參數。統計確定該產品屬性參數和非產品屬性參數中的至少一個高頻詞，並利用確定的至少一個高頻詞為該子集合命名。

本實施例二中的資訊處理設備還可以包括能夠執行實施例一中涉及的各项功能的模組。

通過本申請實施例提供的方法及設備，可以在葉子類目的基礎上，以產品屬性參數為條件劃分細粒度的產品資訊集合，進一步地，還以非產品屬性參數為條件劃分更加細粒度的產品資訊子集合，使得在買家用戶搜索、查詢產品資訊時，以細粒度的產品資訊集合為基礎，可以有效減少搜索、查詢的時間、提高搜索、查詢的準確性以及提高網站伺服器向買家用戶推薦產品資訊的準確度，且細粒度的產品資訊集合也能夠提高對產品資訊進行操作的可用性，減少對產品資訊進行操作時的運算量；並且，在集合和子集合的劃分過程中，對多級類目體系的實質性內容沒有改變，遵守現有的多級類目體系的特點，且充分考慮買家用戶搜索習慣，將買家用戶常用於搜索的產品屬性參數用於劃分集合；另外，對於非標準產品資訊中的產品屬性參數的值不符合設定的條件的各種情況給出了對應的劃分手段，使得各非標準產品資訊能夠盡可能地劃分至相關性高的子集合中；最後，本申請方案還對產品資訊進行去重以及對劃分後的子集合進行命名，有效地減少了對重複產品資訊進行處理所佔用的資源，以及方便管理員根據子集合的名稱對子集合進行管理和以子集合的名稱為關鍵字查找子集合內的產品資訊。

本領域內的技術人員應明白，本申請的實施例可提供為方法、系統、或電腦程式產品。因此，本申請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，本申請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媒體（包括但不限於磁盤記憶體、CD-ROM、光學記憶體等）上實施的電腦程式產品的形式。

本申請是參照根據本申請實施例的方法、設備（系統）、和電腦程式產品的流程圖和／或方框圖來描述的。應理解可由電腦程式指令實現流程圖和／或方框圖中的每一流程和／或方框、以及流程圖和／或方框圖中的流程和／或方框的結合。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編程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編程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／或方框圖一個方框或多個方框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編程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／或方框圖一個方框或多個方框中指定的功能。

這些電腦程式指令也可裝載到電腦或其他可編程資料處理設備上，使得在電腦或其他可編程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編程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／或方框圖一個方框或多個方框中指定的功能的步驟。

儘管已描述了本申請的優選實施例，但本領域內的技術人員一旦得知了基本創造性概念，則可對這些實施例做出另外的變更和修改。所以，所附申請專利範圍意欲解釋為包括優選實施例以及落入本申請範圍的所有變更和修改。

顯然，本領域的技術人員可以對本申請進行各種改動和變型而不脫離本申請的精神和範圍。這樣，倘若本申請的這些修改和變型屬於本申請之申請專利範圍及其等同技術的範圍之內，則本申請也意圖包含這些改動和變型在內。

【圖式簡單說明】

圖 1 為本申請中的資訊處理設備在多級類目體系下的示意圖；

圖 2 為本申請實施例一中資訊處理方法示意圖；

圖 3 為本申請實施例二中資訊處理設備結構示意圖；

圖 4 為本申請實施例二中資訊處理設備中的標準參數確定模組結構示意圖。

【主要元件符號說明】

11：標準參數確定模組

12：第一劃分模組

13：非標準參數確定模組

14：相似度確定模組

15：第二劃分模組

16：判斷模組

17：去重模組

18：命名模組

21：屬性參數確定子模組

22：次數確定子模組

23：選擇子模組

七、申請專利範圍：

104年7月17日修正本

1.一種資訊處理的方法，其特徵在於，該方法包括：

確定葉子類目下各標準產品資訊中的至少一個產品屬性參數，其中此步驟具體包括：

確定葉子類目下各標準產品資訊中的全部產品屬性參數；

確定設定時長內每一產品屬性參數作為搜索條件，在該葉子類目下進行產品資訊搜索的次數；

從該全部產品屬性參數中選擇至少一個產品屬性參數，

其中，選擇的任一產品屬性參數作為搜索條件進行產品資訊搜索的次數達到第一閾值且該產品屬性參數的值為離散型；

利用確定的該至少一個產品屬性參數對該標準產品資訊進行劃分，得到該葉子類目下的多個集合；

其中，劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同。

2.如申請專利範圍第 1 項所述的方法，其中，對該標準產品資訊進行劃分，得到該葉子類目下的多個集合之後，該方法還包括：

針對劃分後得到的每一集合，確定集合中各標準產品資訊中的至少一個非產品屬性參數；

確定同一集合內各標準產品資訊的至少一個非產品屬性參數的相似度；

按照確定的相似度對同一集合內各標準產品資訊做進一步劃分，得到該集合下的多個子集合；

其中，劃分在同一子集合內的任意兩個標準產品資訊的該至少一個非產品屬性參數之間的相似度達到第二閾值。

3.如申請專利範圍第 2 項所述的方法，其中，該方法還包括：

確定葉子類目下的非標準產品資訊，該非標準產品資訊中的用於劃分集合的至少一個產品屬性參數的值不符合設定的條件；

得到集合下的多個子集合之後，該方法還包括：

判斷非標準產品資訊中的用於劃分集合的全部產品屬性參數的值是否都不符合設定的條件；

若是，則確定該非標準產品資訊中的至少一個非產品屬性參數，並根據該非產品屬性參數將該非標準產品資訊劃分至一個子集合內，其中，該非標準產品資訊的非產品屬性參數與劃分至的子集合內的標準產品資訊中的非產品屬性參數的相似度達到第三閾值；

若否，則確定非標準產品資訊中用於劃分集合的產品屬性參數中取值符合設定的條件的產品屬性參數，並確定包含取值符合設定的條件的產品屬性參數的集合，以及，根據非標準產品資訊的非產品屬性參數，將該非標準產品資訊劃分至確定的集合內的一個子集合中，其中，該非標準產品資訊的非產品屬性參數與劃分至的子集合內的標準

產品資訊中的非產品屬性參數的相似度最高。

4.如申請專利範圍第 3 項所述的方法，其中，在非標準產品資訊中的用於劃分集合的全部產品屬性參數的值不符合設定的條件，且該非標準產品資訊的非產品屬性參數與任一子集合內的標準產品資訊中的非產品屬性參數的相似度低於第三閾值時，將該非標準產品資訊劃分至特定集合。

5.如申請專利範圍第 3 或 4 項所述的方法，其中，該方法還包括：

確定重複的標準產品資訊，保留其中一個標準產品資訊，去除剩餘的標準產品資訊；

確定重複的非標準產品資訊，保留其中一個非標準產品資訊，去除剩餘的非標準產品資訊；

該重複的標準產品資訊間的產品屬性參數和非產品屬性參數都相同，該重複的非標準產品資訊間的產品屬性參數和非產品屬性參數都相同。

6.如申請專利範圍第 3 或 4 項所述的方法，其中，

確定每一子集合的標準產品資訊和非標準產品資訊中的產品屬性參數和非產品屬性參數；

統計確定該產品屬性參數和非產品屬性參數中的至少一個高頻詞；

利用確定的至少一個高頻詞為該子集合命名。

7.如申請專利範圍第 1 項所述的方法，其中，得到該葉子類目下的多個集合之後，該方法還包括：

根據得到的各集合中的標準產品資訊進行搜索或產品資訊推薦。

8.一種資訊處理的設備，其特徵在於，該設備包括：

標準參數確定模組，用於確定葉子類目下各標準產品資訊中的至少一個產品屬性參數，其中該標準參數確定模組，包括：

屬性參數確定子模組，用於確定葉子類目下各標準產品資訊中的全部產品屬性參數；

次數確定子模組，確定設定時長內每一產品屬性參數作為搜索條件，在該葉子類目下進行產品資訊搜索的次數；

選擇子模組，用於從該全部產品屬性參數中選擇至少一個產品屬性參數，其中，選擇的任一產品屬性參數作為搜索條件進行產品資訊搜索的次數達到第一閾值且該產品屬性參數的值為離散型；

第一劃分模組，用於利用確定的該至少一個產品屬性參數對該標準產品資訊進行劃分，得到該葉子類目下的多個集合，其中，劃分在同一集合內的標準產品資訊中的該至少一個產品屬性參數的值都相同。

9.如申請專利範圍第 8 項所述的設備，其中，該設備還包括：

非標準參數確定模組，用於針對劃分後得到的每一集合，確定集合中各標準產品資訊中的至少一個非產品屬性參數；

相似度確定模組，用於確定同一集合內各標準產品資訊的至少一個非產品屬性參數的相似度；

第二劃分模組，用於按照確定的相似度對同一集合內各標準產品資訊做進一步劃分，得到該集合下的多個子集合，其中，劃分在同一子集合內的任意兩個標準產品資訊的該至少一個非產品屬性參數之間的相似度達到第二閾值。

10.如申請專利範圍第 9 項所述的設備，其中，該設備還包括：

判斷模組，用於判斷非標準產品資訊中的用於劃分集合的全部產品屬性參數的值是否都不符合設定的條件，若是，則觸發非標準參數確定模組，否則，觸發標準參數確定模組，該非標準產品資訊中的用於劃分集合的至少一個產品屬性參數的值不符合設定的條件；

該非標準參數確定模組，還用於確定該非標準產品資訊中的至少一個非產品屬性參數；

該相似度確定模組，用於確定非標準產品資訊的非產品屬性參數與各子集合內的標準產品資訊中的非產品屬性參數的相似度；

該標準參數確定模組，還用於確定非標準產品資訊中用於劃分集合的產品屬性參數中符合設定的條件的產品屬性參數；

該第一劃分模組，還用於確定包含該符合設定的條件的產品屬性參數的集合；

該第二劃分模組，還用於根據非產品屬性參數，將該非標準產品資訊劃分至一個子集合內，其中，該非標準產品資訊的非產品屬性參數與劃分至的子集合內的標準產品資訊中的非產品屬性參數的相似度達到第三閾值，或者，

根據非標準產品資訊的非產品屬性參數，將該非標準產品資訊劃分至第一劃分模組確定的集合內的一個子集合中，其中，該非標準產品資訊的非產品屬性參數與劃分至的子集合內的標準產品資訊中的非產品屬性參數的相似度最高。

11.如申請專利範圍第 10 項所述的設備，其中，

該第一劃分模組，還用於在非標準產品資訊中的用於劃分集合的全部產品屬性參數的值不符合設定的條件，且該非標準產品資訊的非產品屬性參數與任一子集合內的標準產品資訊中的非產品屬性參數的相似度低於第三閾值時，將該非標準產品資訊劃分至特定集合。

12.如申請專利範圍第 10 或 11 項所述的設備，其中，該設備還包括：

去重模組，用於確定重複的標準產品資訊，保留其中一個標準產品資訊，去除剩餘的標準產品資訊，以及，確定重複的非標準產品資訊，保留其中一個非標準產品資訊，去除剩餘的非標準產品資訊，該重複的標準產品資訊間的產品屬性參數和非產品屬性參數都相同，該重複的非標準產品資訊間的產品屬性參數和非產品屬性參數都相同。

13. 如申請專利範圍第 10 或 11 項所述的設備，其中，該設備還包括：

命名模組，用於確定每一子集合的標準產品資訊和非標準產品資訊中的產品屬性參數和非產品屬性參數，統計確定該產品屬性參數和非產品屬性參數中的至少一個高頻詞，並利用確定的至少一個高頻詞為該子集合命名。

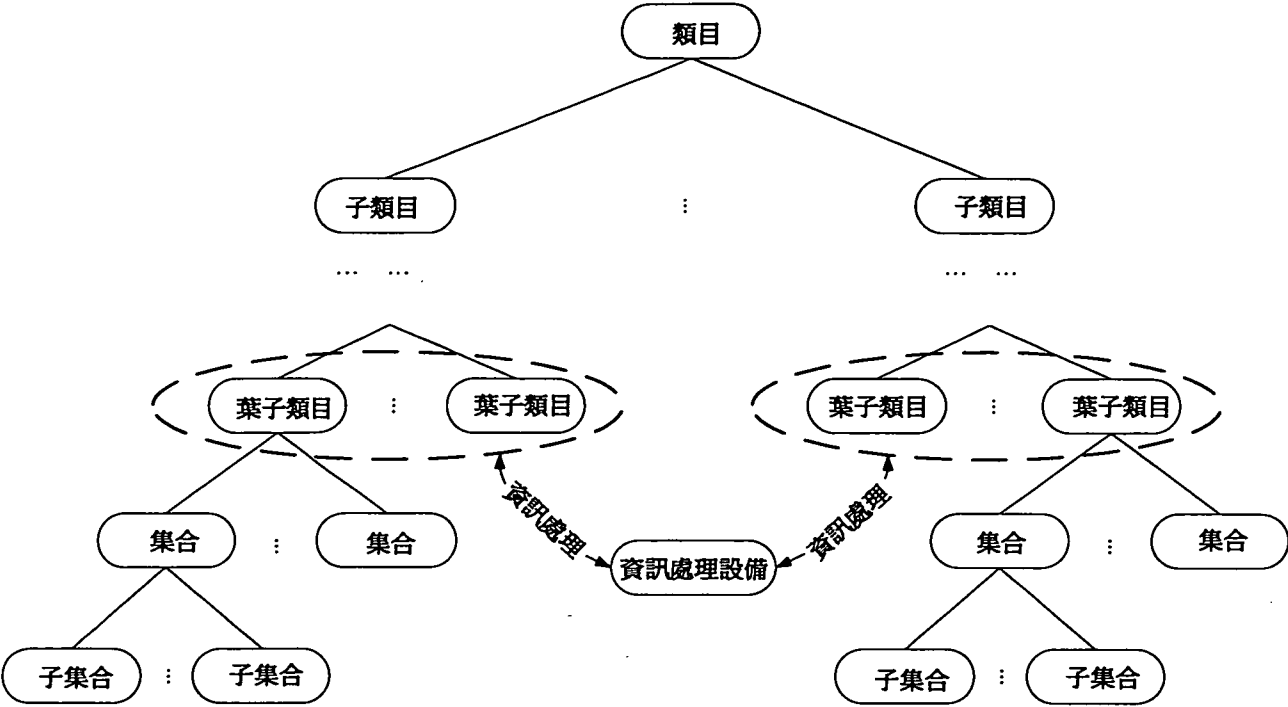


圖 1

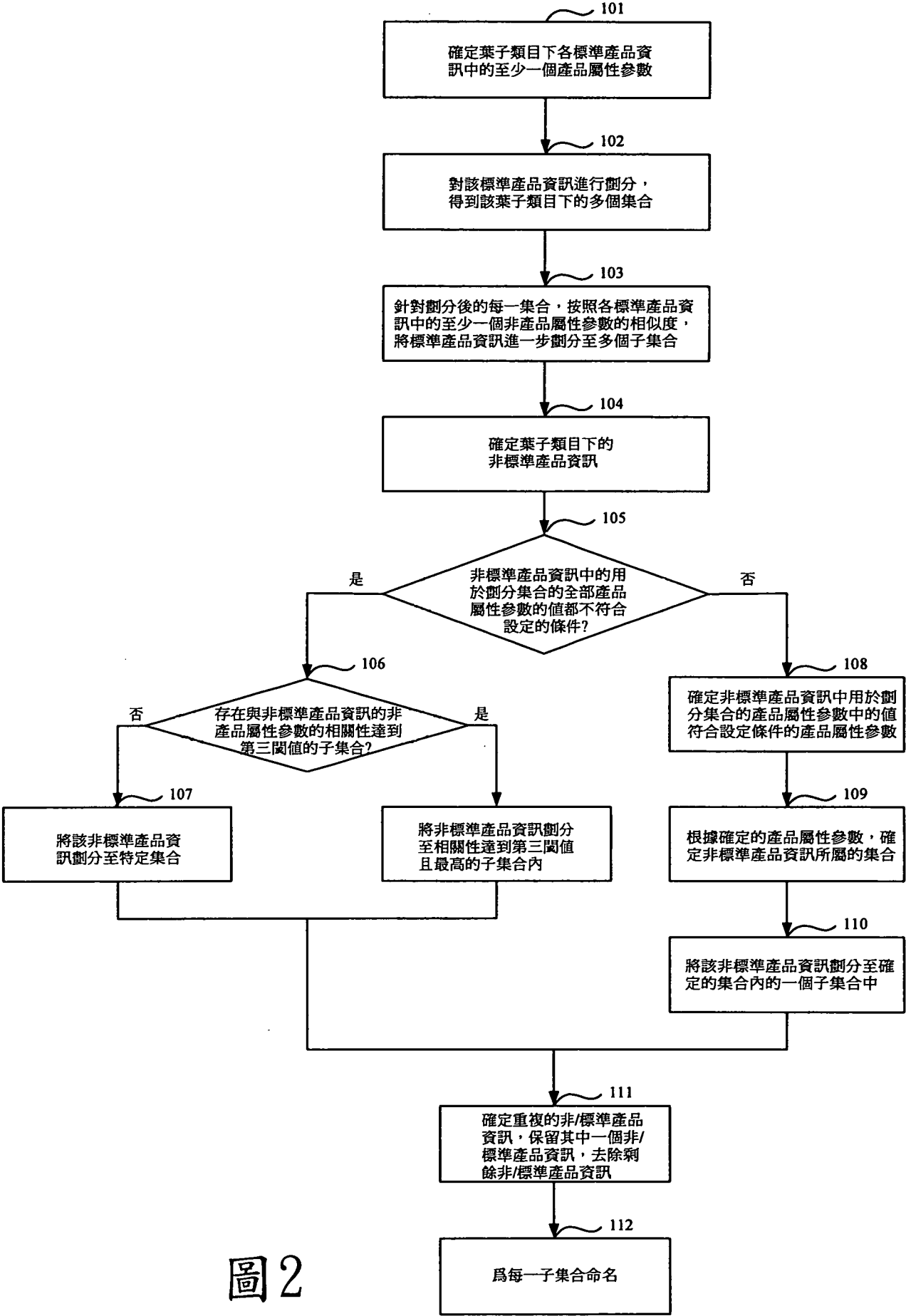


圖2

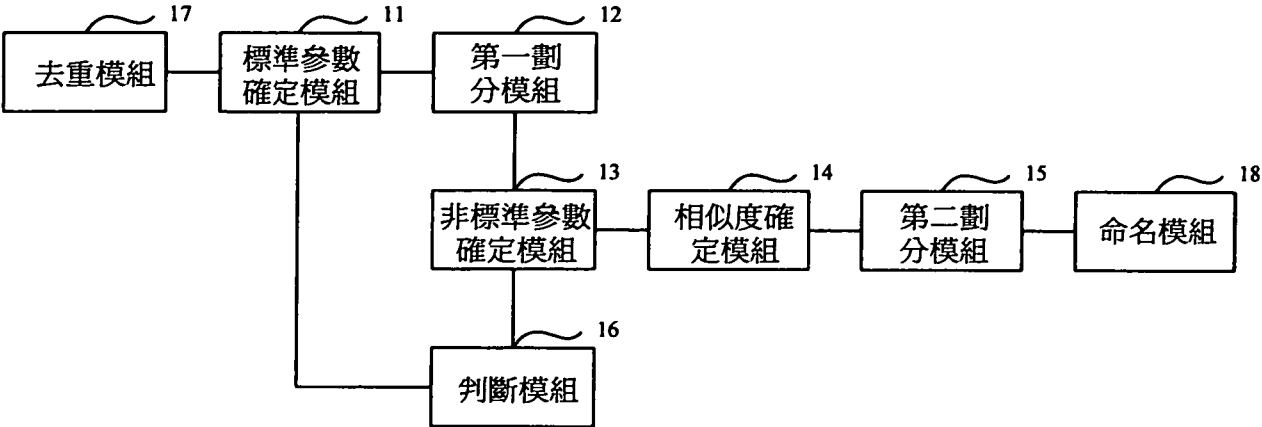


圖 3

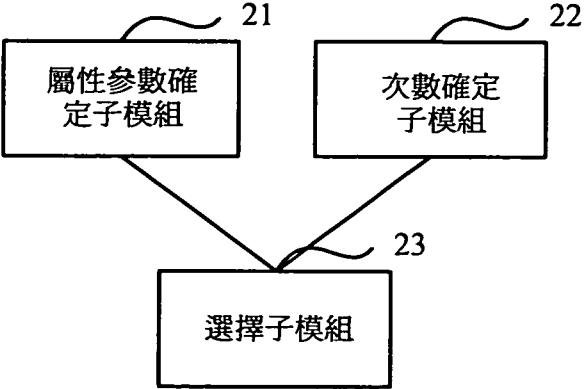


圖 4