



(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 202016758 A

(43) 公開日：中華民國 109 (2020) 年 05 月 01 日

(21) 申請案號：107137786

(22) 申請日：中華民國 107 (2018) 年 10 月 25 日

(51) Int. Cl.：

*G06F16/00 (2019.01)**G06Q30/02 (2012.01)*(71) 申請人：財團法人資訊工業策進會(中華民國) INSTITUTE FOR INFORMATION INDUSTRY  
(TW)

臺北市大安區和平東路 2 段 106 號 11 樓

(72) 發明人：郭欣逸 KUO, HSIN YI (TW)；楊又權 YANG, YOU CYUAN (TW)；黃文發 HUANG, WEN FA (TW)；王文男 WANG, WEN NAN (TW)；陳棟易 CHEN, PING I (TW)

(74) 代理人：陳翠華

申請實體審查：有 申請專利範圍項數：20 項 圖式數：5 共 29 頁

(54) 名稱

知識圖譜產生裝置、方法及其電腦程式產品

(57) 摘要

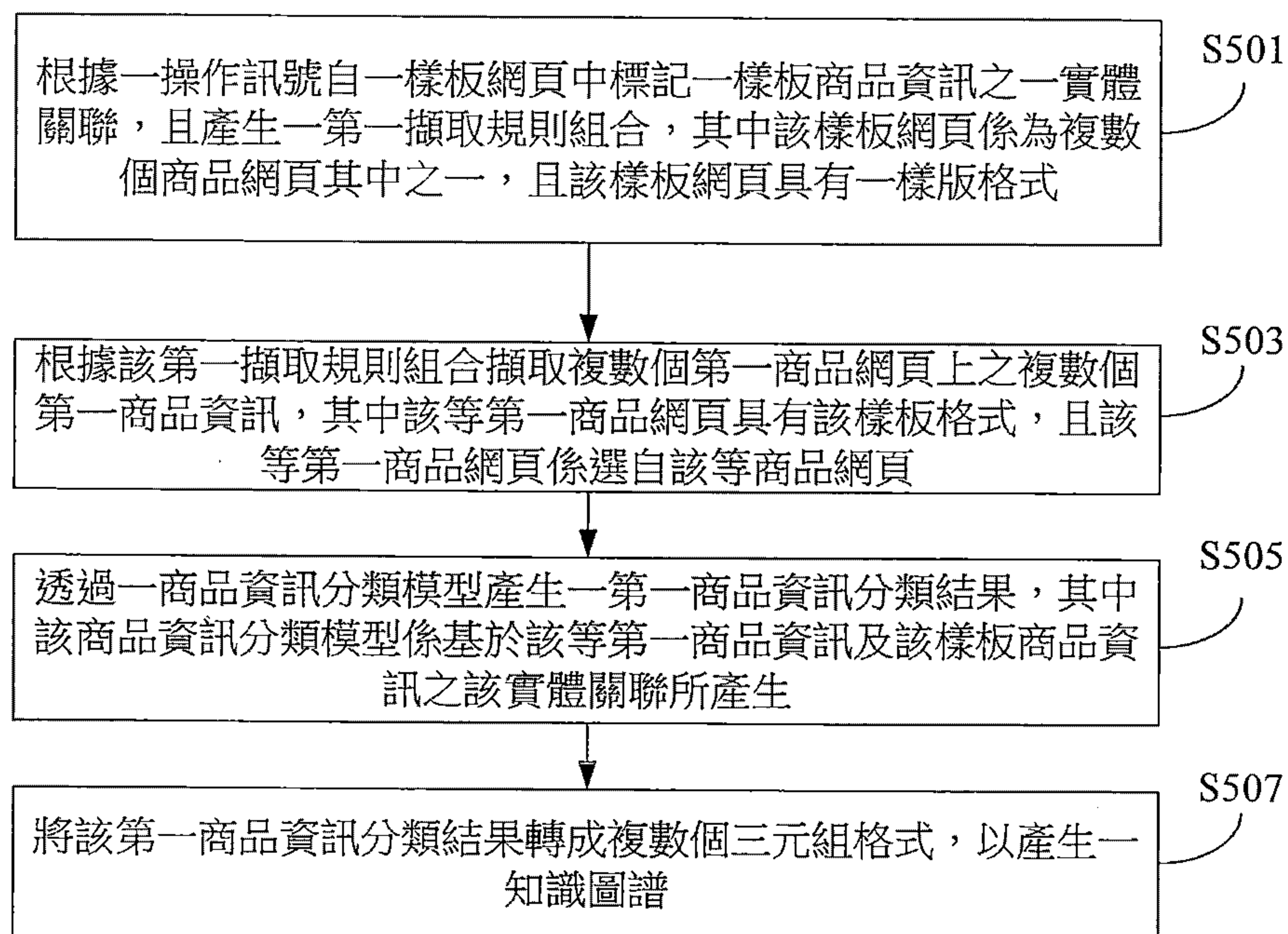
一種知識圖譜產生裝置、方法及其電腦程式產品被揭露。該裝置根據操作訊號自樣板網頁中標記樣板商品資訊之實體關聯且產生擷取規則組合，樣板網頁係為複數個商品網頁其中之一，且樣板網頁具有樣版格式。該裝置根據擷取規則組合擷取複數個第一商品網頁之複數個第一商品資訊，該等第一商品網頁具有該樣板格式，且該等第一商品網頁係選自該等商品網頁。該裝置透過商品資訊分類模型產生商品資訊分類結果，商品資訊分類模型係基於該等第一商品資訊及樣板商品資訊之實體關聯所產生。該裝置將商品資訊分類結果轉成複數個三元組格式，以產生知識圖譜。

A knowledge graph generating apparatus, method and computer program product thereof are provided. The apparatus marks an entity-relationship of the template product information in the template webpage and generates a capture rule set according to the operating signal, wherein the template webpage is one of multiple product webpages, and the template webpage has a pattern format. The apparatus extracts a plurality of first product information of the first product webpages according to the capture rule set, wherein the first product webpages have the template format, and the first product webpages are selected from the product webpages. The apparatus generates a product information classification result through a product information classification model, wherein the product information classification model is generated based on the entity-relationship of the first product information and the entity-relationship of the template product information. The apparatus converts the product information classification result into several semantic triples to generate a knowledge graph.

指定代表圖：

符號簡單說明：

S501~S507 . . . 步驟



第 5 圖

## 發明摘要

### 【發明名稱】(中文/英文)

知識圖譜產生裝置、方法及其電腦程式產品/ KNOWLEDGE GRAPH GENERATEING APPARATUS, METHOD, AND COMPUTER PROGRAM PRODUCT THEREOF

### 【中文】

一種知識圖譜產生裝置、方法及其電腦程式產品被揭露。該裝置根據操作訊號自樣板網頁中標記樣板商品資訊之實體關聯且產生擷取規則組合，樣板網頁係為複數個商品網頁其中之一，且樣板網頁具有樣版格式。該裝置根據擷取規則組合擷取複數個第一商品網頁之複數個第一商品資訊，該等第一商品網頁具有該樣板格式，且該等第一商品網頁係選自該等商品網頁。該裝置透過商品資訊分類模型產生商品資訊分類結果，商品資訊分類模型係基於該等第一商品資訊及樣板商品資訊之實體關聯所產生。該裝置將商品資訊分類結果轉成複數個三元組格式，以產生知識圖譜。

### 【英文】

A knowledge graph generating apparatus, method and computer program product thereof are provided. The apparatus marks an entity-relationship of the template product information in the template webpage and generates a capture rule set according to the operating signal, wherein the template webpage is one of multiple product webpages, and the template webpage has a pattern format. The apparatus extracts a plurality of first product information of the first product webpages according to the capture rule set, wherein the first product webpages have the template format, and the first product webpages are selected from the

product webpages. The apparatus generates a product information classification result through a product information classification model, wherein the product information classification model is generated based on the entity-relationship of the first product information and the entity-relationship of the template product information. The apparatus converts the product information classification result into several semantic triples to generate a knowledge graph.

**【代表圖】**

**【本案指定代表圖】**：第（5）圖。

**【本代表圖之符號簡單說明】**：

S501~S507：步驟

**【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】(中文/英文)

知識圖譜產生裝置、方法及其電腦程式產品/ KNOWLEDGE GRAPH  
GENERATEING APPARATUS, METHOD, AND COMPUTER PROGRAM  
PRODUCT THEREOF

## 【技術領域】

【0001】 本發明係關於一種知識圖譜產生裝置、方法及其電腦程式產品。具體而言，本發明係關於一種基於樣版格式將商品網頁資訊自動轉為知識圖譜之一知識圖譜產生裝置、方法及其電腦程式產品。

## 【先前技術】

【0002】 隨著人工智慧 (artificial intelligence) 的快速發展，智慧問答系統 (例如：聊天機器人 (chat bot)) 等相關服務已逐漸地應用於各種領域，透過模擬人類對話，而達到依據使用者的需求提供各式服務 (例如：資訊獲取、回答產品相關之問題、推薦符合需求之產品等等)。

【0003】 這些智慧問答系統皆需要事先建置背景知識庫 (例如：問答語料庫或知識圖譜 (Knowledge Graph))，以達到快速並準確的回應相關訊息。然而，建置知識圖譜相當耗時費工，一般廠商或企業雖擁有龐大的商品網頁資訊 (通常具有半結構化形式)，卻缺乏一個有效率的方法將其轉變為具結構化的形式，使得商品網頁資料難以生成相應之知識圖譜。在此情況下，商品網頁資訊的內容便無法有效率的應用，或是作為智慧問答系統的基礎資料。

【0004】 有鑑於此，如何有效率地基於網頁資訊產生相應之知識圖

譜，乃業界亟需努力之目標。

### 【發明內容】

【0005】 為解決前述問題，本發明提供了一種知識圖譜產生裝置、方法及其電腦程式產品。

【0006】 本發明所提供之知識圖譜產生裝置包含一網路介面、一操作介面及一處理器，且該處理器電性連接至該網路介面及該操作介面。該網路介面，用以連結至複數個商品網頁。該操作介面，用以產生一操作訊號。該處理器執行下列運作：(a)根據該操作訊號自一樣板網頁中標記一樣板商品資訊之一實體關聯 (Entity-relationship)，且產生一第一擷取規則組合，其中該樣板網頁係為該複數個商品網頁其中之一，且該樣板網頁具有一樣版格式，(b)透過該網路介面連結至複數個第一商品網頁，並根據該第一擷取規則組合擷取該等第一商品網頁上之複數個第一商品資訊，其中該等第一商品網頁具有該樣版格式，且該等第一商品網頁係選自該等商品網頁，(c)透過一商品資訊分類模型產生一第一商品資訊分類結果，其中該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生，以及(d)將該第一商品資訊分類結果轉成複數個三元組 (Semantic triple) 格式，以產生一知識圖譜。

【0007】 本發明所提供之一種知識圖譜產生方法適用於一知識圖譜產生裝置，用以產生一操作訊號。該知識圖譜產生方法由該知識圖譜裝置所執行且包含下列步驟：(a)根據該操作訊號自一樣板網頁中標記一樣板商品資訊之一實體關聯，且產生一第一擷取規則組合，其中該樣板網頁係為該等商品網頁其中之一，且該樣板網頁具有一樣版格式，(b)根據該第一擷取規

則組合擷取複數個第一商品網頁上之複數個第一商品資訊，其中該等第一商品網頁具有該樣板格式，且該等第一商品網頁係選自該等商品網頁，(c)透過一商品資訊分類模型產生一第一商品資訊分類結果，其中該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生，以及(d)將該第一商品資訊分類結果轉成複數個三元組格式，以產生一知識圖譜。

**【0008】** 本發明所提供之電腦程式產品經由一電子裝置載入該電腦程式產品後，該電子裝置執行該電腦程式產品所包含之複數個程式指令，以執行一知識圖譜產生方法。該知識圖譜產生方法包含下列步驟：(a)根據一操作訊號自一樣板網頁中標記一樣板商品資訊之一實體關聯，且產生一第一擷取規則組合，其中該樣板網頁係為複數個商品網頁其中之一，且該樣板網頁具有一樣版格式，(b)根據該第一擷取規則組合擷取複數個第一商品網頁上之複數個第一商品資訊，其中該等第一商品網頁具有該樣板格式，且該等第一商品網頁係選自該等商品網頁，(c)透過一商品資訊分類模型產生一第一商品資訊分類結果，其中該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生，以及(d)將該第一商品資訊分類結果轉成複數個三元組格式，以產生一知識圖譜。

**【0009】** 由上述說明可知，本發明提供之知識圖譜產生裝置、方法及其電腦程式產品，藉由分析具有一樣板格式之樣板網頁，產生基於該樣板格式之擷取規則組合。更進一步時可由於複數張其他商品網頁同樣具有該樣板格式，本發明可透過該擷取規則組合自動擷取其他商品網頁之商品資訊。另外，由於本發明更將該樣板網頁之樣板商品資訊標記一實體關聯，因此可

根據一商品資訊分類模型（基於多個商品資訊之實體關聯及該樣板商品資訊之實體關聯產生）產生商品資訊分類結果。最後，本發明透過將該商品資訊分類結果轉成三元組格式，產生知識圖譜。此外，本發明亦可根據多個樣板格式建置擷取規則模型，且當商品網頁更新時，進一步更新該知識圖譜。由於降低建置知識圖譜的人力耗費，藉由本發明提供之技術可有效率的產生多個網頁之知識圖譜，作為智慧問答系統的基礎資料，解決了習知建置知識圖譜耗時費工的問題。

**【0010】** 以下結合圖式闡述本發明之詳細技術及實施方式，俾使本發明所屬技術領域中具有通常知識者能理解所請求保護之發明之特徵。

#### **【圖式簡單說明】**

**【0011】** 第1圖係描繪第一實施方式之知識圖譜產生裝置之架構示意圖；

**【0012】** 第2A圖係描繪第一實施方式中樣板網頁之示意圖；

**【0013】** 第2B圖係描繪第一實施方式中樣板網頁上資訊之實體關聯示意圖；

**【0014】** 第3圖係描繪樣板網頁之網頁原始碼示意圖；

**【0015】** 第4A圖係描繪多張商品網頁之示意圖；

**【0016】** 第4B圖係描繪根據第4A圖之商品網頁所產生之知識圖譜之示意圖；以及

**【0017】** 第5圖係描繪第二實施方式之知識圖譜產生方法之流程圖。

#### **【實施方式】**

**【0018】** 以下將透過實施方式來解釋本發明所提供之知識圖譜產生

裝置、方法及其電腦程式產品。然而，該等實施方式並非用以限制本發明需在如該等實施方式所述之任何環境、應用或方式方能實施。因此，關於實施方式之說明僅為闡釋本發明之目的，而非用以限制本發明之範圍。應理解，在以下實施方式及圖式中，與本發明非直接相關之元件已省略而未繪示。

**【0019】** 本發明之第一實施方式為一知識圖譜產生裝置1，其架構示意圖係描繪於第1圖。知識圖譜產生裝置1包含一網路介面11、一操作介面13及一處理器15，處理器15電性連接至網路介面11及操作介面13。網路介面11為一連接至網路之介面或本發明所屬技術領域中具有通常知識者所知悉之其他可與網路連接之介面。操作介面13為一輸入介面或本發明所屬技術領域中具有通常知識者所知悉之其他能接收輸入訊號並產生操作訊號之介面。處理器13可為各種處理單元、中央處理單元（Central Processing Unit；CPU）、微處理器或本發明所屬技術領域中具有通常知識者所知悉之其他計算裝置。需說明者，於本發明之其他實施方式中，操作介面13可被設置為外在於知識圖譜產生裝置1。

**【0020】** 於本實施方式中，處理器15先將具有一樣板格式之一樣版網頁（例如：自複數個商品網頁中選擇之一網頁）作為一標準，基於該樣板網頁之該樣板格式產生一擷取規則組合。由於該等商品網頁同樣具有該樣板格式，該擷取規則組合將用於後續自動擷取其他商品網頁之商品資訊（容後說明）。另外，處理器15更將該樣板網頁之樣板商品資訊分別對應至一實體關聯（Entity-relationship），並將該樣板商品資訊之實體關聯用於後續標記其他商品網頁之商品資訊（容後說明）。

**【0021】** 具體而言，處理器15接收來自操作介面13之操作訊號133，

根據操作訊號133自一樣板網頁中標記一樣板商品資訊之實體關聯，且產生一第一擷取規則組合。該樣板網頁係為該複數個商品網頁其中之一，且該樣板網頁具有一樣版格式。須說明者，本發明所述之樣板格式為一種半結構化資訊，但本發明未限制樣版格式須採用哪一種半結構化的形式。有關半結構化資訊的樣板格式，例如：一網頁表格或是一層疊樣式表（Cascading Style Sheets；CSS）或者是以列舉方式說明商品及產品相關資訊之網頁內容型態，都屬於本發明保護的範疇。於某些實施方式中，該實體關聯至少包含該商品之一實體（entity）、一關聯（relation）及對應至該關聯之至少一屬性值（property）。另外，於某些實施方式中，該第一擷取規則組合與該樣板格式之至少一類別名稱相關聯。

**【0022】** 為便於理解，謹先詳述本實施方式如何自一樣板網頁中標記樣板商品資訊之實體關聯（如何產生一第一擷取規則組合容後說明）。請參第2A圖、第2B圖及第3圖所示之一具體範例，但其非用以限制本發明之範圍。於該具體範例中，一拍賣網站有複數個鞋子的商品網頁，且該等商品網頁具有同一樣板格式（如第3圖所示，以CSS格式設計之樣板網頁21之網頁原始碼301）。

**【0023】** 如第2A圖所示，樣板網頁21為該等商品網頁其中之一，並顯示了有關一X鞋款之資訊。首先，使用者可透過操作介面13將樣板網頁21上之X鞋款之資訊分類為實體、關聯或屬性值，操作介面13可依據分類後的資訊產生一操作訊號133，傳送給處理器15，其中操作訊號133載有前述的分類結果。如第2B圖之實體關聯示意圖所示，使用者將樣板網頁21上的資訊「X鞋款」分類為一實體，「適合性別」分類為一關聯，「男」分類為對應「適合

性別」關聯的一屬性值。類似的，對應該「X鞋款」實體，使用者根據樣板網頁21上之資訊，將「類型」分類為一關聯，「路跑鞋」分類為對應「類型」關聯的一屬性值。對於樣板網頁21上之其他資訊，使用者亦可一一進行分類，茲不贅言。接著，處理器15根據接收到的操作訊號133，將樣板網頁21上的樣板商品資訊標記出前述之實體關聯。須說明者，於某些實施方式中，該知識圖譜產生裝置1亦可由外部接收一操作訊號來標記樣板商品資訊之實體關聯。

**【0024】** 以下說明處理器15如何根據操作訊號133產生擷取規則組合。第3圖為以CSS格式設計之樣板網頁21之網頁原始碼301，其顯示各欄位所對應之類別（class）名稱。處理器15接收到操作介面13所傳來的操作訊號133後，處理器15在原始碼301中找出至少一類別名稱及其所對應之標記內容，藉此產生第一擷取規則組合。舉例而言，如第3圖虛線框起來的部分，使用者將類別名稱為「equipment\_basic\_information eq-basic-info」的區塊標記為「商品規格」，將類別名稱為「basic\_info\_item basic-info-item」的區塊標記為「商品規格說明」。須說明者，此處僅用以例示，但其非用以限制本發明之範圍，使用者亦可根據不同之需求或是類別來標記網頁的不同區塊。由於該等商品網頁同樣具有該樣板格式，後續處理該等商品網頁時，處理器15即可根據類別名稱擷取該等商品網頁上對應之資訊，並根據標記的內容，將擷取下來的商品內容分類（例如：該等擷取之內容屬於「商品規格」或是「商品規格說明」等）。

**【0025】** 隨後，處理器15透過網路介面11連結至其他商品網頁，由於其他商品網頁同樣具有該樣板格式，處理器15即可根據該第一擷取規則組

合，擷取其他商品網頁上之複數個第一商品資訊。接著，處理器15透過一商品資訊分類模型產生一第一商品資訊分類結果，該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生。須說明者，本發明未限制該商品資訊分類模型須由知識圖譜產生裝置1本身產生，亦可自外部接收一建置好之商品資訊分類模型。

**【0026】** 具體而言，該商品資訊分類模型可根據下列運作產生該第一商品資訊分類結果。首先，由處理器（可為內部處理器15或一外部處理器）基於前述該等商品網頁上擷取下來的商品資訊，進行一斷詞分析，接著產生一斷詞結果。須說明者，該斷詞分析可包含一詞彙相似度分析，可透過如BM25 (Best Match 25)、TF/IDF (Term frequency-inverse document frequency) 等演算法實現，其更可包含去雜訊、斷句、斷詞等運作。隨後，根據樣板商品資訊之該實體關聯以及該斷詞結果，進行一關聯性分析，透過該關聯性分析將斷詞結果標記實體關聯。最後，根據標記完實體關聯的斷詞結果，產生該第一商品資訊分類結果。需說明者，有關斷詞分析的細節、關聯性分析的內容為何以及各斷詞結果如何分類並非本發明之重點，本發明所屬技術領域中具有通常知識者應可理解其內容，故不贅言。

**【0027】** 須說明者，由於不同之商品網頁資訊用語可能不一致(例如：足型、容量、尺寸等)，該商品資訊分類模型於訓練階段時，可透過大量已標記實體關聯之同類產品(例如：擷取自該等第一商品網頁之商品資訊)進行訓練，並根據已生成之商品資訊分類結果反饋其結果進行訓練及整合，以處理資訊用語不一致的問題。如此一來，透過該商品資訊分類模型之關聯性分析，可提高標記實體關聯的準確性。本發明所屬技術領域中具有通常知識

者應可理解如何訓練商品資訊分類模型，故不贅言。

**【0028】** 最後，處理器15將該第一商品資訊分類結果轉成複數個三元組格式（亦即Triplestore，以實體、關聯、屬性值順序呈現之資料），以產生一知識圖譜。請參考第4B圖之一知識圖譜4之示意圖，其例示了根據第4A圖之商品網頁上之商品資訊所產生之一知識圖譜。須說明者，第4B圖僅用以例示知識圖譜內容之實體關聯，但其非用以限制本發明之範圍，以習知方式儲存及各種態樣表示的知識圖譜都屬於本發明保護的範疇，本發明所屬技術領域中具有通常知識者應可理解其內容，故不贅言。此外，一商品網頁可能根據資訊的不同具有多張知識圖譜，本發明所屬技術領域中具有通常知識者應可理解產生多張知識圖譜之方式及其內容，故不贅言。

**【0029】** 於某些實施方式中，處理器15更根據該第一擷取規則組合擷取一第二商品網頁上之一第二商品資訊，其中該第二商品網頁具有該樣板格式。於該等實施方式中，可採用第一實施方式所述之技術，透過該商品資訊分類模型產生一第二商品資訊分類結果，其中該第二商品資訊分類結果指示該等第二商品資訊之該實體關聯。最後，根據該第二商品資訊分類結果，更新該知識圖譜。

**【0030】** 於某些實施方式中，處理器15更可根據多張具有不同樣板格式之樣板網頁（例如：來自不同拍賣網站之商品網頁），產生一擷取規則模型。具體而言，處理器15根據一第二操作訊號自一第二樣板網頁中標記一第二商品之實體關聯，且產生一第二擷取規則組合，其中該第二樣板網頁係為該複數個商品網頁其中之一，且該等第二樣板網頁具有一第二樣板格式。處理器15根據該第一擷取規則組合及該第二擷取規則組合，產生一擷取規則

模型。因此，當處理器15接收到具有不同樣板格式之商品網頁時，即可根據該擷取規則模型，選擇適合之擷取規則組合，擷取該等商品網頁之商品資訊。

【0031】 此外，於某些實施方式中，當該等第一商品網頁有更新時，處理器15根據該第一擷取規則組合，重新擷取該等第一商品網頁以取得更新之第一商品資訊。接著，處理器15透過該商品資訊分類模型產生更新之第一商品資訊分類結果。隨後，處理器15根據該更新之第一商品資訊分類結果，更新該知識圖譜。

【0032】 由上述說明可知，知識圖譜產生裝置1藉由分析具有一樣板格式之樣板網頁，產生基於該樣板格式之擷取規則組合。對於也具有該樣板格式之複數張其他商品網頁，知識圖譜產生裝置1可透過該擷取規則組合自動擷取其他商品網頁之商品資訊。另外，由於知識圖譜產生裝置1更將該樣板網頁之樣板商品資訊標記一實體關聯，知識圖譜產生裝置1可根據一商品資訊分類模型（基於多個商品資訊之實體關聯及該樣板商品資訊之實體關聯產生）產生商品資訊分類結果。最後，知識圖譜產生裝置1透過將該商品資訊分類結果轉成三元組格式，產生知識圖譜。此外，知識圖譜產生裝置1亦可根據多個樣板格式建置擷取規則模型，且當商品網頁更新時，進一步更新該知識圖譜。由於降低建置知識圖譜的人力耗費，藉由本發明提供之技術可有效率的產生多個網頁之知識圖譜，作為智慧問答系統的基礎資料，解決了習知建置知識圖譜耗時費工的問題。

【0033】 本發明之第二實施方式為一種知識圖譜產生方法，其流程圖係描繪於第5圖。知識圖譜產生方法適用於一電子裝置，例如：第一實施方

式所述之知識圖譜產生裝置1。知識圖譜產生方法透過步驟S501至步驟S507產生知識圖譜。

**【0034】** 首先，於步驟S501，由該電子裝置根據一操作訊號自一樣板網頁中標記一樣板商品資訊之一實體關聯，且產生一第一擷取規則組合，其中該樣板網頁係為複數個商品網頁其中之一，且該樣板網頁具有一樣版格式。須說明者，本發明所述之樣板格式為一半結構化資訊，本發明未限制樣版格式須採用哪一種半結構化的形式。有關半結構化資訊的樣板格式，例如：一網頁表格或是一層疊樣式表（Cascading Style Sheets；CSS），都屬於本發明保護的範疇。於某些實施方式中，該實體關聯至少包含該商品之一實體（entity）、一關聯（relation）及對應至該關聯之至少一屬性值（property）。另外，於某些實施方式中，該第一擷取規則組合與該樣板格式之至少一類別名稱相關聯。

**【0035】** 接著，於步驟S503，由該電子裝置根據該第一擷取規則組合擷取複數個第一商品網頁上之複數個第一商品資訊，其中該等第一商品網頁具有該樣板格式，且該等第一商品網頁係選自該等商品網頁。

**【0036】** 隨後，於步驟S505，由該電子裝置透過一商品資訊分類模型產生一第一商品資訊分類結果，其中該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生。於某些實施方式中，步驟S505可包含下列步驟：基於該等第一商品資訊，進行一斷詞分析以產生一斷詞結果；以及根據該樣板商品資訊之該實體關聯與該斷詞結果，進行一關聯性分析以產生該第一商品資訊分類結果，其中該第一商品資訊分類結果指示該等第一商品資訊之該實體關聯。

【0037】 之後，於步驟S507，由該電子裝置將該第一商品資訊分類結果轉成複數個三元組格式，以產生一知識圖譜。須說明者，一商品網頁可能根據資訊的不同具有多張知識圖譜，本發明所屬技術領域中具有通常知識者應可理解產生多張知識圖譜之方式及其內容，故不贅言。

【0038】 於某些實施方式中，該知識圖譜產生方法更包含下列步驟：根據該第一擷取規則組合擷取一第二商品網頁上之一第二商品資訊，其中該第二商品網頁具有該樣板格式。透過該商品資訊分類模型產生一第二商品資訊分類結果，其中該第二商品資訊分類結果指示該等第二商品資訊之該實體關聯。根據該第二商品資訊分類結果，更新該知識圖譜。

【0039】 於某些實施方式中，該知識圖譜產生方法更包含下列步驟：根據一第二操作訊號自一第二樣板網頁中標記一第二商品之實體關聯，以產生一第二擷取規則組合，其中該第二樣板網頁係為該等商品網頁其中之一，且該等第二樣板網頁具有一第二樣板格式。接著，該知識圖譜產生方法更包含一步驟以根據該第一擷取規則組合及該第二擷取規則組合，產生一擷取規則模型。

【0040】 於某些實施方式中，該知識圖譜產生方法更包含下列步驟：當該等第一商品網頁有更新時，根據該第一擷取規則組合重新擷取該等第一商品網頁以取得更新之第一商品資訊。透過該商品資訊分類模型產生更新之第一商品資訊分類結果。根據該更新之第一商品資訊分類結果，更新該知識圖譜。

【0041】 除了上述步驟，第二實施方式亦能執行第一實施方式所描述之知識圖譜產生裝置1之所有運作及步驟，具有同樣之功能，且達到同樣之

技術效果。本發明所屬技術領域中具有通常知識者可直接瞭解第二實施方式如何基於上述第一實施方式以執行此等運作及步驟，具有同樣之功能，並達到同樣之技術效果，故不贅述。

**【0042】** 第二實施方式所闡述之知識圖譜產生方法可由具有複數個指令之一電腦程式產品實現。各電腦程式產品可為能被於網路上傳輸之檔案，亦可被儲存於一非暫態電腦可讀取儲存媒體中。針對各電腦程式產品，在其所包含之該等指令被載入至一電子裝置（例如：知識圖譜產生裝置1）之後，該電腦程式產品執行第二實施方式所述之知識圖譜產生方法。該非暫態電腦可讀取儲存媒體可為一電子產品，例如：一唯讀記憶體（read only memory；ROM）、一快閃記憶體、一軟碟、一硬碟、一光碟（compact disk；CD）、一隨身碟、一可由網路存取之資料庫或本發明所屬技術領域中具有通常知識者所知且具有相同功能之任何其他儲存媒體。

**【0043】** 需說明者，於本發明專利說明書及申請專利範圍中，某些用語（包含：擷取規則組合、商品網頁、商品資訊、商品資訊分類結果、樣板網頁及樣板格式）前被冠以「第一」或「第二」，該等「第一」及「第二」僅用來區分不同之用語。

**【0044】** 綜上所述，本發明所提供之知識圖譜產生技術（至少包含裝置、方法及其電腦程式產品），藉由分析具有一樣板格式之樣板網頁，產生基於該樣板格式之擷取規則。由於複數張其他商品網頁同樣具有該樣板格式，本發明所提供之知識圖譜產生技術可透過該擷取規則自動擷取其他商品網頁之商品資訊。另外，由於本發明所提供之知識圖譜產生技術更將該樣板網頁之樣板商品資訊標記一實體關聯，即可根據一商品資訊分類模型（基

於多個商品資訊之實體關聯及該樣板商品資訊之實體關聯產生) 產生商品資訊分類結果。最後，本發明所提供之知識圖譜產生技術透過將該商品資訊分類結果轉成三元組格式，產生知識圖譜。此外，本發明所提供之知識圖譜產生技術亦可根據多個樣板格式建置擷取規則模型，且當商品網頁更新時，進一步更新該知識圖譜。由於降低建置知識圖譜的人力耗費，藉由本發明所提供之知識圖譜產生技術可有效率的產生多個網頁之知識圖譜，作為智慧問答系統的基礎資料，解決了習知建置知識圖譜耗時費工的問題。

**【0045】** 上述實施方式僅用來例舉本發明之部分實施態樣，以及闡釋本發明之技術特徵，而非用來限制本發明之保護範疇及範圍。任何本發明所屬技術領域中具有通常知識者可輕易完成之改變或均等性之安排均屬於本發明所主張之範圍，而本發明之權利保護範圍以申請專利範圍為準。

#### **【符號說明】**

##### **【0046】**

1：知識圖譜產生裝置

11：網路介面

13：操作介面

15：處理器

133：操作訊號

21：樣板網頁

301：網頁原始碼

4：知識圖譜

S501-S507：步驟

**【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. 一種知識圖譜(Knowledge Graph)產生裝置，包含：
  - 一網路介面，用以連結至複數個商品網頁；
  - 一操作介面，用以產生一操作訊號；以及
  - 一處理器，電性連接至該網路介面及該操作介面，且執行下列運作：
    - (a)根據該操作訊號自一樣板網頁中標記一樣板商品資訊之一實體關聯(Entity-relationship)，且產生一第一擷取規則組合，其中該樣板網頁係為該複數個商品網頁其中之一，且該樣板網頁具有一樣版格式；
    - (b)透過該網路介面連結至複數個第一商品網頁，並根據該第一擷取規則組合擷取該等第一商品網頁上之複數個第一商品資訊，其中該等第一商品網頁具有該樣版格式，且該等第一商品網頁係選自該等商品網頁；
    - (c)透過一商品資訊分類模型產生一第一商品資訊分類結果，其中該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生；以及
    - (d)將該第一商品資訊分類結果轉成複數個三元組 ( Semantic triple ) 格式，以產生一知識圖譜。
2. 如請求項1所述之知識圖譜產生裝置，其中該樣板格式為一個網頁表格或一層疊樣式表 ( Cascading Style Sheets ; CSS )。
3. 如請求項1所述之知識圖譜產生裝置，其中該第一擷取規則組合與該樣板格式之至少一類別名稱相關聯。
4. 如請求項1所述之知識圖譜產生裝置，其中該實體關聯至少包含該商品之

- 一實體(entity)、一關聯(relation)及對應至該關聯之至少一屬性值(property)。
5. 如請求項1所述之知識圖譜產生裝置，其中該商品資訊分類模型包含下列運作：
- 基於該等第一商品資訊，進行一斷詞分析以產生一斷詞結果；以及
- 根據該樣板商品資訊之該實體關聯與該斷詞結果，進行一關聯性分析以產生該第一商品資訊分類結果，其中該第一商品資訊分類結果指示該等第一商品資訊之該實體關聯。
6. 如請求項1所述之知識圖譜產生裝置，其中該處理器更執行下列運作：
- 根據該第一擷取規則組合擷取一第二商品網頁上之一第二商品資訊，其中該第二商品網頁具有該樣板格式；
- 透過該商品資訊分類模型產生一第二商品資訊分類結果，其中該第二商品資訊分類結果指示該等第二商品資訊之該實體關聯；以及
- 根據該第二商品資訊分類結果，更新該知識圖譜。
7. 如請求項1所述之知識圖譜產生裝置，該處理器更執行下列運作：
- 根據一第二操作訊號自一第二樣板網頁中標記一第二商品之實體關聯(Entity-relationship)，且產生一第二擷取規則組合，其中該第二樣板網頁係為該複數個商品網頁其中之一，且該等第二樣板網頁具有一第二樣板格式；以及
- 根據該第一擷取規則組合及該第二擷取規則組合，產生一擷取規則模型。
8. 如請求項1所述之知識圖譜產生裝置，該處理器更執行下列運作：

當該等第一商品網頁有更新時，根據該第一擷取規則組合重新擷取該等第一商品網頁以取得更新之第一商品資訊；

透過該商品資訊分類模型產生更新之第一商品資訊分類結果；以及根據該更新之第一商品資訊分類結果，更新該知識圖譜。

9. 一種知識圖譜產生方法，適用於一知識圖譜產生裝置，用以產生一操作訊號，該知識圖譜產生方法由該知識圖譜產生裝置所執行且包含下列步驟：

(a)根據該操作訊號自一樣板網頁中標記一樣板商品資訊之一實體關聯，且產生一第一擷取規則組合，其中該樣板網頁係為該等商品網頁其中之一，且該樣板網頁具有一樣版格式；

(b)根據該第一擷取規則組合擷取複數個第一商品網頁上之複數個第一商品資訊，其中該等第一商品網頁具有該樣版格式，且該等第一商品網頁係選自該等商品網頁；

(c)透過一商品資訊分類模型產生一第一商品資訊分類結果，其中該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生；以及

(d)將該第一商品資訊分類結果轉成複數個三元組格式，以產生一知識圖譜。

10.如請求項9所述之知識圖譜產生方法，其中該樣版格式為一個網頁表格或一層疊樣式表（Cascading Style Sheets；CSS）。

11.如請求項9所述之知識圖譜產生方法，其中該第一擷取規則組合與該樣版格式之至少一類別名稱相關聯。

12.如請求項9所述之知識圖譜產生方法，其中該實體關聯至少包含該商品之一實體(entity)、一關聯(relation)及對應至該關聯之至少一屬性值(property)。

13.如請求項9所述之知識圖譜產生方法，其中透過該商品資訊分類模型產生該第一商品資訊分類結果之步驟包含下列步驟：

基於該等第一商品資訊，進行一斷詞分析以產生一斷詞結果；以及根據該樣板商品資訊之該實體關聯與該斷詞結果，進行一關聯性分析以產生該第一商品資訊分類結果，其中該第一商品資訊分類結果指示該等第一商品資訊之該實體關聯。

14.如請求項9所述之知識圖譜產生方法，更包含下列步驟：

根據該第一擷取規則組合擷取一第二商品網頁上之一第二商品資訊，其中該第二商品網頁具有該樣板格式；

透過該商品資訊分類模型產生一第二商品資訊分類結果，其中該第二商品資訊分類結果指示該等第二商品資訊之該實體關聯；以及

根據該第二商品資訊分類結果，更新該知識圖譜。

15.如請求項9所述之知識圖譜產生方法，更包含下列步驟：

根據一第二操作訊號自一第二樣板網頁中標記一第二商品之實體關聯(Entity-relationship)，以產生一第二擷取規則組合，其中該第二樣板網頁係為該等商品網頁其中之一，且該等第二樣板網頁具有一第二樣板格式；以及

根據該第一擷取規則組合及該第二擷取規則組合，產生一擷取規則模型。

16.如請求項9所述之知識圖譜產生方法，更包含下列步驟：

當該等第一商品網頁有更新時，根據該第一擷取規則組合重新擷取該等第一商品網頁以取得更新之第一商品資訊；

透過該商品資訊分類模型產生更新之第一商品資訊分類結果；以及  
根據該更新之第一商品資訊分類結果，更新該知識圖譜。

17.一種電腦程式產品，經由一電子裝置載入該電腦程式產品後，該電子裝置執行該電腦程式產品所包含之複數個程式指令，以執行一知識圖譜產生方法，該知識圖譜產生方法包含下列步驟：

(a)根據一操作訊號自一樣板網頁中標記一樣板商品資訊之一實體關聯，且產生一第一擷取規則組合，其中該樣板網頁係為複數個商品網頁其中之一，且該樣板網頁具有一樣版格式；

(b)根據該第一擷取規則組合擷取複數個第一商品網頁上之複數個第一商品資訊，其中該等第一商品網頁具有該樣版格式，且該等第一商品網頁係選自該等商品網頁；

(c)透過一商品資訊分類模型產生一第一商品資訊分類結果，其中該商品資訊分類模型係基於該等第一商品資訊及該樣板商品資訊之該實體關聯所產生；以及

(d)將該第一商品資訊分類結果轉成複數個三元組格式，以產生一知識圖譜。

18.如請求項17所述之電腦程式產品，其中透過該商品資訊分類模型產生該第一商品資訊分類結果之步驟包含下列步驟：

基於該等第一商品資訊，進行一斷詞分析以產生一斷詞結果；以及

根據該樣板商品資訊之該實體關聯與該斷詞結果，進行一關聯性分析以產生該第一商品資訊分類結果，其中該第一商品資訊分類結果指示該等第一商品資訊之該實體關聯。

19.如請求項17所述之電腦程式產品，其中該知識圖譜產生方法更包含下列步驟：

根據該第一擷取規則組合擷取一第二商品網頁上之一第二商品資訊，其中該第二商品網頁具有該樣板格式；

透過該商品資訊分類模型產生一第二商品資訊分類結果，其中該第二商品資訊分類結果指示該等第二商品資訊之該實體關聯；以及

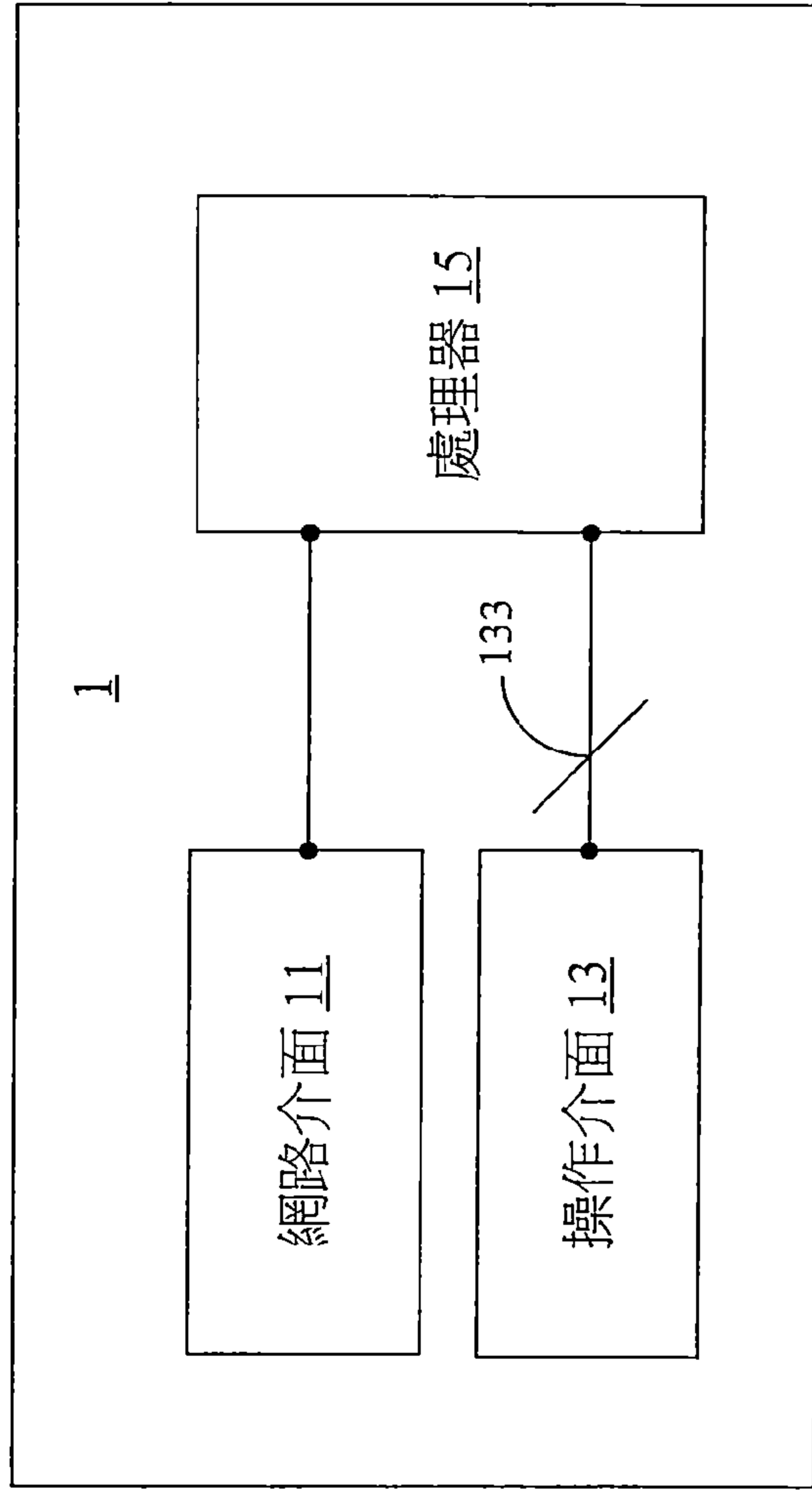
根據該第二商品資訊分類結果，更新該知識圖譜。

20.如請求項17所述之電腦程式產品，其中該知識圖譜產生方法更包含下列步驟：

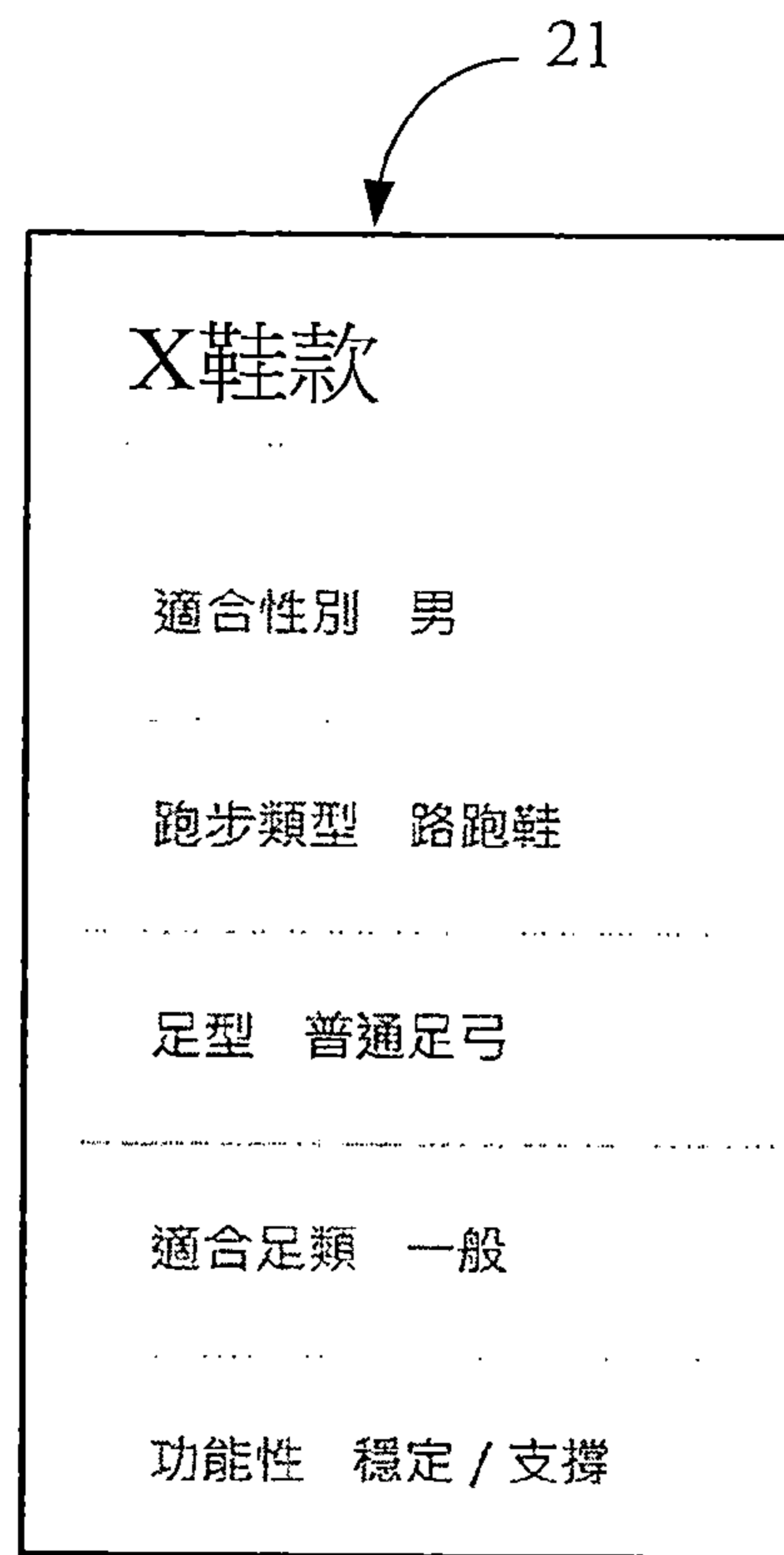
當該等第一商品網頁有更新時，根據該第一擷取規則組合重新擷取該等第一商品網頁以取得更新之第一商品資訊；

透過該商品資訊分類模型產生更新之第一商品資訊分類結果；以及  
根據該更新之第一商品資訊分類結果，更新該知識圖譜。

圖式



第 1 圖



第 2A 圖

實體	關聯	屬性值
X鞋款	適合性別	男
X鞋款	跑步類型	路跑鞋
X鞋款	足型	普通足弓
X鞋款	適合足類	一般
X鞋款	功能性	穩定/支撐

第 2B 圖

301

```

    <div class="sub_main_topic equipment-title"><h3>規格</h3></div>
  <div id='equipment_info' class="main_content equipment-box">
    <div class='information_box'>
      <div class="equipment_basic_information eq-basic-info">
        <div class="basic_info_item basic-info-item">
          <div class="basic_info_title eq-info-til">適合性別</div>
          <div class="basic_info_content basic-info-type">男</div>
        </div>
        <div class="basic_info_item basic-info-item">
          <div class="basic_info_title eq-info-til">跑步類型</div>
          <div class="basic_info_content basic-info-type">路跑鞋</div>
        </div>
        <div class="basic_info_item basic-info-item">
          <div class="basic_info_title eq-info-til">足型</div>
          <div class="basic_info_content basic-info-type">普通足弓</div>
        </div>
      </div>
    </div>
  </div>

```

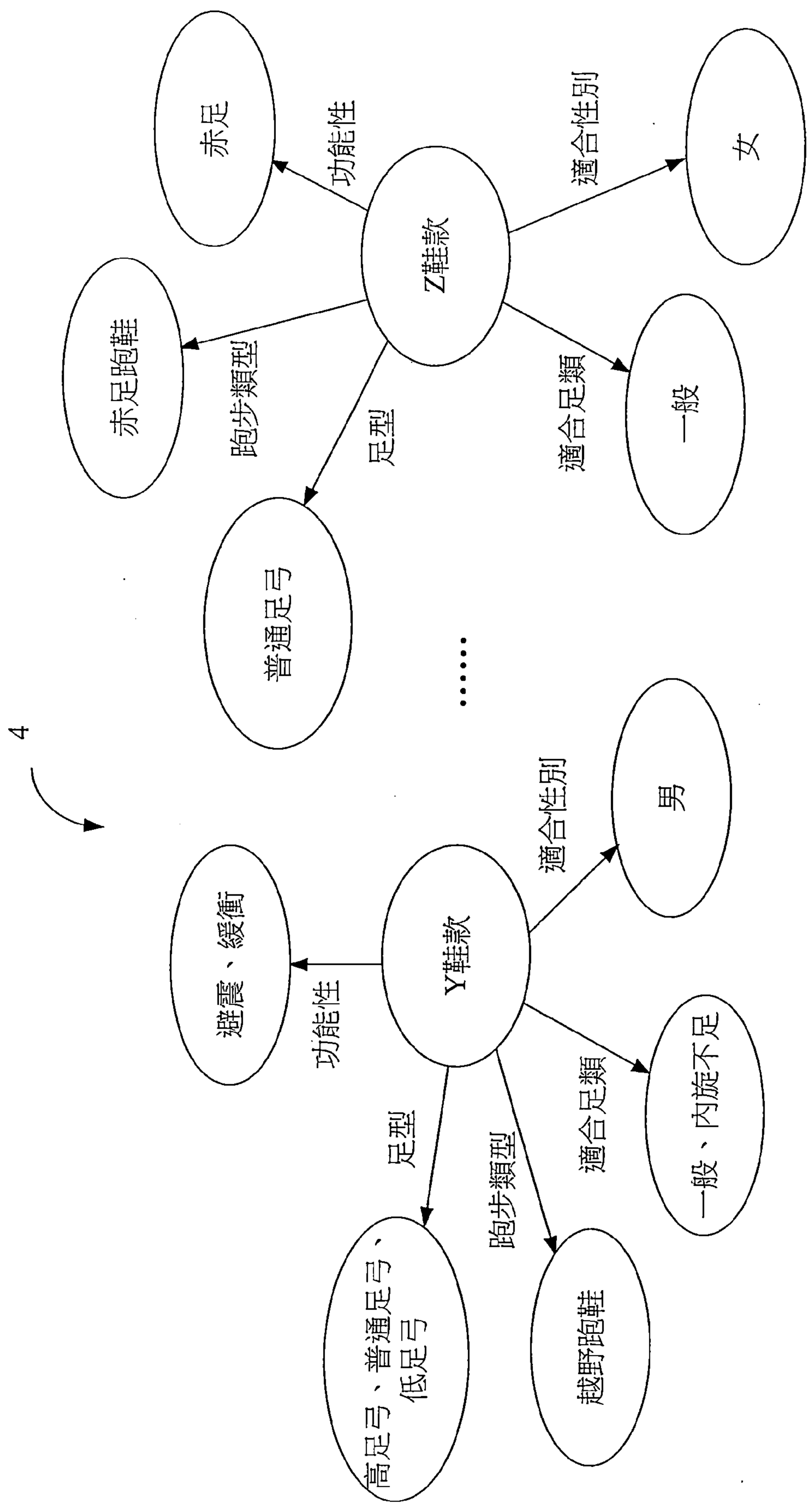
第 3 圖

Y鞋款	
適合性別	男
跑步類型	越野跑鞋
足型	高足弓、普通足弓、低足弓
適合足類	一般、內旋不足
功能性	避震 / 緩衝

.....

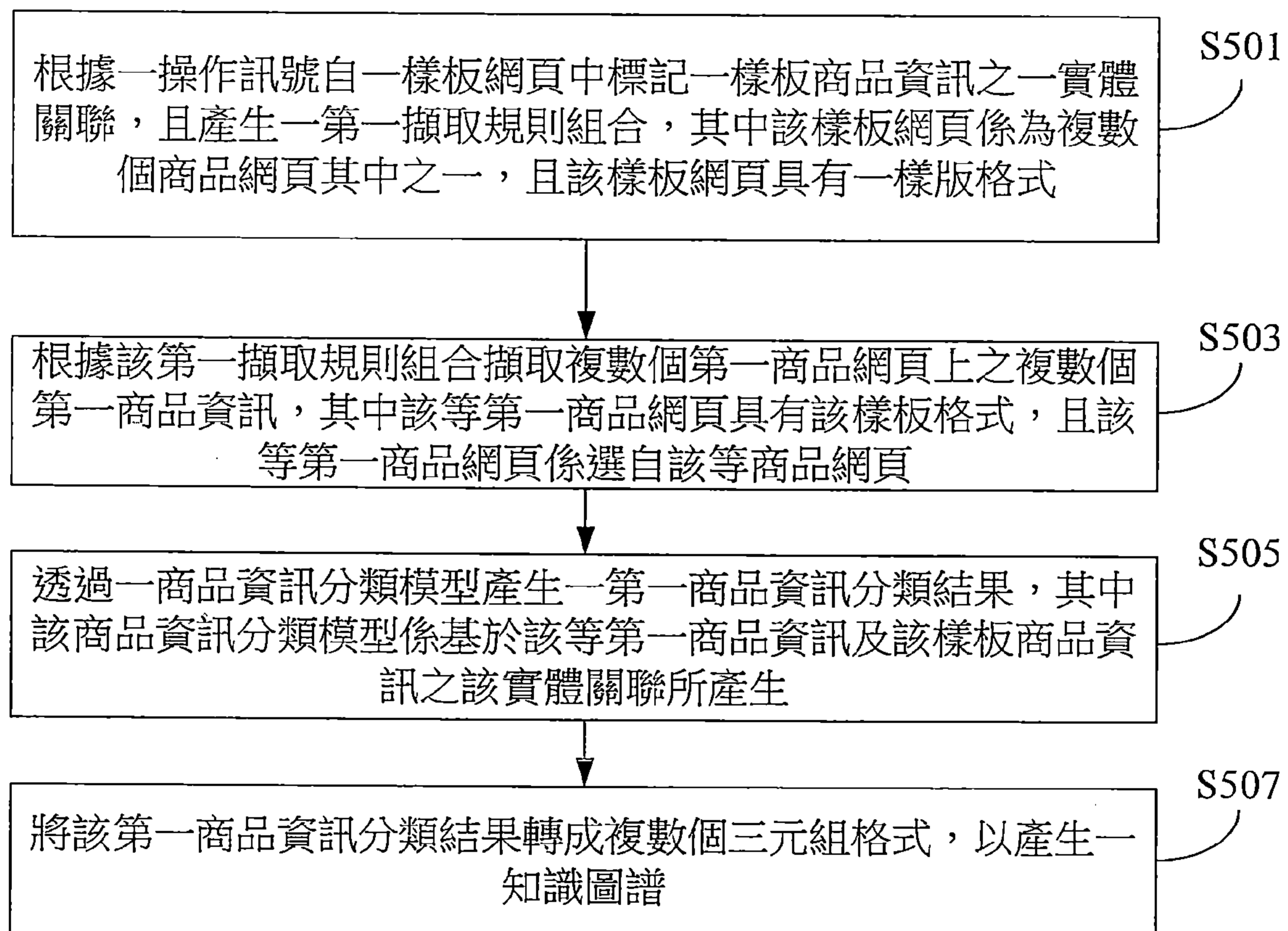
Z鞋款	
適合性別	女
跑步類型	赤足跑鞋
足型	普通足弓
適合足類	一般
功能性	赤足

第4A圖



4

第4B圖



第 5 圖