

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2014 年 12 月 31 日(31.12.2014)



(10) 国際公開番号
WO 2014/207941 A1

- (51) 国際特許分類:
G06F 17/21 (2006.01)
- (21) 国際出願番号: PCT/JP2013/067942
- (22) 国際出願日: 2013 年 6 月 28 日(28.06.2013)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 楽天株式会社(RAKUTEN, INC.) [JP/JP]; 〒1400002 東京都品川区東品川四丁目 1 2 番 3 号 Tokyo (JP).
- (72) 発明者: 落合 由紀子(OCHIAI, Yukiko); 〒1400002 東京都品川区東品川四丁目 1 2 番 3 号 楽天株式会社内 Tokyo (JP).
- (74) 代理人: 特許業務法人はるか国際特許事務所 (HARUKA PATENT & TRADEMARK ATTORNEYS); 〒1600023 東京都新宿区西新宿三丁目 1 番 4 号 ウエル新都心ビル 4 階 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[続葉有]

(54) Title: DETERMINATION DEVICE, DETERMINATION METHOD, AND PROGRAM

(54) 発明の名称: 判定装置、判定方法、及びプログラム

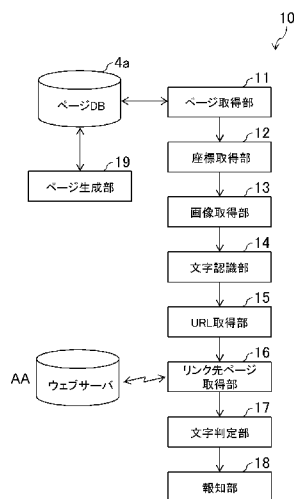


FIG. 6
4a Page database
11 Page acquisition unit
12 Coordinate acquisition unit
13 Image acquisition unit
14 Text recognition unit
15 URL acquisition unit
16 Linked-page acquisition unit
17 Text determination unit
18 Notification unit
19 Page generation unit
AA Web server

(57) Abstract: This determination device (10) is provided with the following: an image acquisition unit (13) that acquires an image from a link area with which a URL is associated; a linked-page acquisition unit (16) that acquires, from a storing means for storing content, the linked page specified by the URL associated with the aforementioned link area; and a text determination unit (17) that determines, on the basis of the image acquired by the image acquisition unit (13) and the linked page acquired by the linked-page acquisition unit (16), whether or not the association between the link area and the URL is correct.

(57) 要約: 判定装置 (10) は、URL が対応付けられたリンクエリア内の画像を取得する画像取得部 (13) と、コンテンツを記憶する記憶手段から、前記リンクエリアに対応付けられた前記 URL によって特定されたリンク先ページを取得するリンク先ページ取得部 (16) と、画像取得部 (13) により取得された画像と、リンク先ページ取得部 (16) により取得されたリンク先ページとに基づいて、前記リンクエリアと前記 URL との対応付けの正誤を判定する文字判定部 (17) と、を含む。



添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称： 判定装置、判定方法、及びプログラム

技術分野

[0001] 本発明は、判定装置、判定方法、及びプログラムに関する。

背景技術

[0002] 近年、ユーザがウェブサイトアクセスしてウェブページに含まれるテキストや写真を選択すると、該テキストや写真にリンクされたURL（Uniform Resource Locator）のウェブページ等のコンテンツを表示するシステムが普及している。例えば、ウェブショッピングで利用されるウェブカタログには複数の商品が一覧表示されており、ユーザが所望の商品のコードや写真を選択すると、該商品の情報を掲載した商品ページ（コンテンツ）が表示される。このようなシステムでは、商品コードや写真がクリックブルエリアに設定されており、各クリックブルエリアには商品ページのURLがリンクされている。

[0003] また、特許文献1には、雑誌、カタログ等の印刷物作成用のデータをPDFデータに変換する際に、テキストや写真等に、リンク情報を付加する技術が記載されている。

先行技術文献

特許文献

[0004] 特許文献1：特開2004-289357号公報

発明の概要

発明が解決しようとする課題

[0005] 上記システムでは、クリックブルエリアとURLとの対応付けを、サイト運営者が手作業で行っているため、対応付けの誤りが生じることがある。またサイト運営者は、上記対応付けの正誤の検査を手作業で行わなければならない手間がかかる。

[0006] 本発明は、上記課題に鑑みてなされたものであり、その目的は、ウェブペ

ージに含まれる情報と、該情報のリンク先のURLとの対応付けの正誤を容易に判定することができる判定装置、判定方法、及びプログラムを提供することにある。

課題を解決するための手段

[0007] 上記課題を解決するために、本発明に係る判定装置は、URLが対応付けられたリンクエリア内又は該リンクエリア周辺から得られるリンク元情報を取得する情報取得手段と、コンテンツを記憶する記憶手段から、前記リンクエリアに対応付けられた前記URLによって特定されたコンテンツを取得するコンテンツ取得手段と、前記情報取得手段により取得された前記リンク元情報と、前記コンテンツ取得手段により取得された前記コンテンツとに基づいて、前記リンクエリアと前記URLとの対応付けの正誤を判定する判定手段と、を含むことを特徴とする。

[0008] また、前記判定装置は、前記情報取得手段により取得された前記リンク元情報を文字として認識する文字認識手段をさらに含んでもよい。また、前記判定手段は、前記文字認識手段により認識された前記文字と、前記コンテンツ取得手段により取得された前記コンテンツとに基づいて、前記リンクエリアと前記URLとの対応付けの正誤を判定してもよい。

[0009] また、前記判定手段は、前記文字認識手段により認識された前記文字が、前記コンテンツ取得手段により取得された前記コンテンツに含まれるか否かにより、前記対応付けの正誤を判定してもよい。

[0010] また、前記判定手段は、前記文字認識手段により認識された複数の前記文字のうち所定割合以上の数の文字が、前記コンテンツ取得手段により取得された前記コンテンツに含まれるか否かにより、前記対応付けの正誤を判定してもよい。

[0011] また、前記判定装置は、ページ内における前記リンクエリアの座標を取得する座標取得手段をさらに含んでもよい。また、前記情報取得手段は、前記座標取得手段により取得された前記リンクエリアの座標に基づいて、前記リンク元情報を取得してもよい。

- [0012] また、前記判定装置は、前記対応付けが誤っている場合にエラーメッセージを報知する報知手段をさらに含んでもよい。
- [0013] また、前記判定装置は、ページ内の前記リンクエリアを設定するとともに、予め前記ＵＲＬが登録されたテーブルから、設定された前記リンクエリアに対応する前記ＵＲＬを取得し、前記リンクエリアと前記ＵＲＬとを対応付ける生成手段をさらに含んでもよい。
- [0014] また、前記テーブルにおいて、前記ＵＲＬは、前記リンクエリアの配置順に対応して登録されていてもよい。
- [0015] また、本発明に係る判定方法は、ＵＲＬが対応付けられたリンクエリア内又は該リンクエリア周辺から得られるリンク元情報を取得する情報取得ステップと、コンテンツを記憶する記憶手段から、前記リンクエリアに対応付けられた前記ＵＲＬによって特定されたコンテンツを取得するコンテンツ取得ステップと、前記情報取得ステップにより取得された前記リンク元情報と、前記コンテンツ取得ステップにより取得された前記コンテンツとに基づいて、前記リンクエリアと前記ＵＲＬとの対応付けの正誤を判定する判定ステップと、を含むことを特徴とする。
- [0016] また、本発明に係るプログラムは、ＵＲＬが対応付けられたリンクエリア内又は該リンクエリア周辺から得られるリンク元情報を取得する情報取得手段、コンテンツを記憶する記憶手段から、前記リンクエリアに対応付けられた前記ＵＲＬによって特定されたコンテンツを取得するコンテンツ取得手段、及び、前記情報取得手段により取得された前記リンク元情報と、前記コンテンツ取得手段により取得された前記コンテンツとに基づいて、前記リンクエリアと前記ＵＲＬとの対応付けの正誤を判定する判定手段、としてコンピュータを機能させるためのプログラムである。このプログラムは、ＣＤ－ＲＯＭやＤＶＤ－ＲＯＭなどのコンピュータ可読情報記憶媒体に格納されてもよい。

発明の効果

- [0017] 本発明によれば、リンクエリア内の文字とリンクエリアに対応付けられた

URLに対応するコンテンツとに基づいて、リンクエリアとURLとの対応付けの正誤を判定する。よって、ウェブページに含まれる情報と、該情報のリンク先のURLとの対応付けの正誤を容易に判定することができる。

図面の簡単な説明

- [0018] [図1]ウェブカタログの一例を示す図である。
- [図2]商品ページの一例を示す図である。
- [図3]判定装置のハードウェア構成図である。
- [図4]HTMLファイルの一例を示す図である。
- [図5]クリックブルエリアの一例を示す図である。
- [図6]判定装置の機能ブロック図である。
- [図7]実施例1に係る判定装置の動作フロー図である。
- [図8]商品ページの一例を示す図である。
- [図9]HTMLファイルの一例を示す図である。
- [図10]ウェブカタログ及び商品ページの一例を示す図である。
- [図11]HTMLファイルの一例を示す図である。
- [図12]ページ生成部の動作フロー図である。
- [図13]商品コード一覧のテーブルを示す図である。
- [図14]URL一覧のテーブルを示す図である。
- [図15]HTMLファイルの一例を示す図である。
- [図16]判定装置のハードウェア構成図である。

発明を実施するための形態

- [0019] 以下、本発明の一実施形態について図面に基づき詳細に説明する。以下では、インターネットショッピングのウェブサイトで利用されるウェブカタログを例に挙げる。

- [0020] 図1は、ウェブカタログの一例を示す図である。ユーザがインターネットショッピングのウェブサイトにアクセスすると、ユーザ端末にウェブカタログ（ウェブページ）が表示される。ウェブカタログには、複数の商品、例えば、しょうゆ、リンゴ、パン等の情報が一覧表示されている。各商品の商品

コード、写真、テキスト等には、対応する商品の詳細情報を掲載した商品ページ（コンテンツ）のURLが対応付けられている。例えば、商品「しょうゆ」の商品名、商品コード及び写真には、図2に示す「しょうゆ」の商品ページのURLが対応付けられている。例えばユーザがウェブカタログ（図1参照）において商品「しょうゆ」の写真を選択すると、ユーザ端末には「しょうゆ」の商品ページ（図2参照）が表示される。「しょうゆ」の商品ページには、商品名、写真、商品コード、値段、容量、原材料、メーカー名等の詳細情報が掲載されている。同様に、ウェブカタログの商品「リンゴ」、「パン」には、図2に示す「リンゴ」、「パン」の商品ページが対応付けられている。なお、ユーザが商品の購入を希望する場合は、商品ページに表示された購入ボタンを選択する。

[0021] 本実施形態に係る判定装置は、例えば上記のようなインターネットショッピングの形態において、ウェブカタログ（図1参照）に掲載される商品と、該商品の情報を掲載した商品ページ（図2参照）のURLとの対応付けの正誤を容易に判定することを可能にするものである。なお上記判定処理は、サイト運営者の操作に基づき実行される。

[0022] 図3は、本実施形態に係る判定装置のハードウェア構成図である。同図に示すように、判定装置10は、通信部1、CPU2、メモリ3、及び記憶部4を含んでいる。判定装置10を構成するハードウェア要素はバスにより相互にデータの授受が可能に接続されている。通信部1は、例えばインターネットよりユーザ端末と情報の送受信を行う。CPU2は装置各部を制御したり、各種の情報処理を実行したりする。メモリ3は、各種のプログラムやデータを保持する。またメモリ3には、CPU2の作業領域も確保される。

[0023] 記憶部4には、ページDB4aが含まれる。ページDB4aには、複数のウェブカタログに対応する複数のHTMLファイルが記憶されている。HTMLファイルはサイト運営者により作成される。具体的には、サイト運営者は、ウェブカタログの元となるPDF形式の画像データ（カタログデータ）を例えば印刷会社から入手し、入手したカタログデータに基づいて、クリッ

カブルエリアと、これに対応するリンク先URLとを設定したHTMLファイルを作成する。このHTMLファイルとこのHTMLファイルにおいて指定された画像ファイルとをユーザ端末のウェブブラウザで表示することにより、図1に示すウェブカタログが表示される。なお、クリックカブルエリアの指定とリンク先の指定とは、例えば、HTMLとは別のJavaScript（登録商標）ファイルにて指定してもよい。ウェブカタログに掲載された商品に対応付けられたリンク先の商品ページは、サイト運営者により作成され、ウェブサーバ（記憶手段）にアップロードされている。なお記憶部4は、インターネットを介して通信部1、CPU2及びメモリ3に接続されていてもよい。

[0024] 図4は、図1に示すウェブカタログに対応するHTMLファイルの一例である。図5は、図1に示すウェブカタログにおけるクリックカブルエリア（リンクエリア）の一例である。HTMLファイルには、クリックカブルエリアと、これに対応付けるURLとを設定するための要素（mapデータ）が含まれている。ここでは、座標（x1、x2、x3、x4）の位置の四角形のリンクエリアに「しょうゆ」の商品ページのURL「http://aaa.co.jp/しょうゆ.html」が設定され、座標（y1、y2、y3、y4）の位置の四角形のリンクエリアに「リンゴ」の商品ページのURL「http://aaa.co.jp/リンゴ.html」が設定され、座標（z1、z2、z3、z4）の位置の四角形のリンクエリアに「パン」の商品ページのURL「http://aaa.co.jp/パン.html」が設定されている。それぞれのリンクエリアがクリックカブルエリアに対応する。図5に示すように、各クリックカブルエリアは商品コード欄に対応しており、「しょうゆ」のリンクエリアをクリックカブルエリア1、「リンゴ」のリンクエリアをクリックカブルエリア2、「パン」のリンクエリアをクリックカブルエリア3で示している。クリックカブルエリア4については後述する。

[0025] 図6は、判定装置10の機能ブロック図である。同図に示すように、判定装置10は、ページ取得部11、座標取得部12（座標取得手段）、画像取

得部 13（情報取得手段）、文字認識部 14（文字認識手段）、URL 取得部 15、リンク先ページ取得部 16（コンテンツ取得手段）、文字判定部 17（判定手段）、報知部 18（報知手段）、及びページ生成部 19（生成手段）を含んでいる。これらの要素はメモリ 3 に格納されたプログラムを CPU 2 が実行することにより実現される。このプログラムは CD-ROM、DVD-ROM、メモリカードなどのコンピュータ可読情報記憶媒体から判定装置 10 にインストールされてもよいし、インターネット等の通信ネットワークからダウンロードされてもよい。

[0026] ページ取得部 11 は、ページ DB 4 a からウェブカタログに対応する HTML ファイルを取得する。ページ取得部 11 は、例えば図 4 に示す HTML ファイルを取得する。

[0027] 座標取得部 12 は、ページ取得部 11 により取得された HTML ファイルから、クリックابلエリアの座標を取得する。座標取得部 12 は、例えば、図 4 に示す HTML ファイルの map データから、クリックابلエリア 1 の座標（ x_1 、 x_2 、 x_3 、 x_4 ）、クリックابلエリア 2 の座標（ y_1 、 y_2 、 y_3 、 y_4 ）、又は、クリックابلエリア 3 の座標（ z_1 、 z_2 、 z_3 、 z_4 ）を取得する。

[0028] 画像取得部 13 は、図 1 に示すウェブカタログの画像データ（カタログデータ）において、座標取得部 12 により取得された座標に対応するクリックابلエリア内の画像（エリア画像）（リンク元情報）を取得する。画像取得部 13 は、例えば座標（ x_1 、 x_2 、 x_3 、 x_4 ）に対応するクリックابلエリア 1 のエリア画像「★1234」、座標（ y_1 、 y_2 、 y_3 、 y_4 ）に対応するクリックابلエリア 2 のエリア画像「★2345」、又は、座標（ z_1 、 z_2 、 z_3 、 z_4 ）に対応するクリックابلエリア 3 のエリア画像「★3456」を取得する。なお、カタログデータに所定のマーク（ここでは「★」（星マーク））が付されている場合は、画像取得部 13 は、カタログデータを画像認識して、マークを含む所定のエリアをエリア画像として取得してもよい。画像取得部 13 は、クリックابلエリアがテキスト表示欄（図

5のクリッカブルエリア4)の場合はテキスト表示をエリア画像として取得し、クリッカブルエリアが写真欄の場合は写真をエリア画像として取得する。なお、画像取得部13は、クリッカブルエリア内の画像ではなく、その周辺の画像を取得してもよい。例えば、図5のクリッカブルエリア1に対して、該当商品の掲載エリア内に位置する領域をエリア画像として取得しても良い。該当商品の掲載エリアのサイズと、掲載エリアにおけるクリッカブルエリアの位置が分かっているならば、掲載エリア内の領域を特定できる。

[0029] 文字認識部14は、画像取得部13により取得されたエリア画像について文字認識を行う。文字認識部14は、例えば光学文字認識(OCR)の方式により文字認識を行う。文字認識部14は、クリッカブルエリアが商品コード欄の場合は商品コードを文字認識し、クリッカブルエリアがテキスト表示欄の場合はテキストを文字認識する。なお、文字コードを取得できる場合は文字認識処理を行わなくてもよい。

[0030] URL取得部15は、ページ取得部11により取得されたHTMLファイルから、座標取得部12により取得された座標に対応する商品ページのURLを取得する。URL取得部15は、例えば、図4に示すHTMLファイルのmapデータから、クリッカブルエリア1の座標(x1、x2、x3、x4)に対応するURL「http://aaa.co.jp/しょうゆ.html」、クリッカブルエリア2の座標(y1、y2、y3、y4)に対応するURL「http://aaa.co.jp/リンゴ.html」、又は、クリッカブルエリア3の座標(z1、z2、z3、z4)に対応するURL「http://aaa.co.jp/パン.html」を取得する。

[0031] リンク先ページ取得部16は、ウェブサーバから、URL取得部15により取得されたURLの商品ページを取得する。リンク先ページ取得部16は、例えば、ウェブサーバから、図2に示す、URL「http://aaa.co.jp/しょうゆ.html」の商品ページ、URL「http://aaa.co.jp/リンゴ.html」の商品ページ、又は、URL「http://aaa.co.jp/パン.html」の商品ページを取得する。

[0032] 文字判定部 17 は、リンク先ページ取得部 16 により取得された商品ページに、文字認識部 14 により認識された文字が含まれているか否かを判定する。なお、文字判定部 17 は、上記認識文字と一致する文字が上記商品ページに含まれているか否かを判定してもよいし、上記認識文字が複数の場合に所定割合以上の文字数が上記商品ページに含まれているか否かを判定してもよい。また、クリックابلエリアに写真欄が含まれる場合は、文字判定部 17 は、さらに、画像取得部 13 により取得されたエリア画像と、商品ページに掲載されている写真の画像との類似度を算出し、類似度に基づいて上記判定を行ってもよい。この場合、文字判定部 17 は、画像処理判定部として機能し、例えば画像の特徴点を抽出して比較することにより類似度を算出する。

[0033] 報知部 18 は、文字判定部 17 の判定結果に基づいてメッセージを報知する。具体的には、上記認識文字が上記商品ページに含まれていない場合はリンクの対応付けが誤っているため、報知部 18 はエラーメッセージを報知する。なお、上記認識文字が上記商品ページに含まれている場合は、報知部 18 は、リンクの対応付けが正しい旨のメッセージを報知してもよい。

[0034] ページ生成部 19 は、クリックابلエリアの座標と、商品ページの URL とを対応付けて HTML ファイル（図 4 参照）を生成する処理を行う。ページ生成部 19 により生成された HTML ファイルは、ページ DB 4a に記憶される。

[0035] [実施例 1]

図 7 は、実施例 1 に係る判定装置 10 の動作フロー図である。ここでは、図 1 に示すウェブカタログと、図 2 に示す商品ページを例に挙げて説明する。

[0036] サイト運営者は図 1 に示すウェブカタログに対応する HTML ファイル（図 4 参照）を作成し、作成された HTML ファイルがページ DB 4a に記憶されているものとする。また、ウェブサーバには、サイト運営者が作成した、図 2 に示す複数の商品ページがアップロードされているものとする。また

、図1に示すウェブカタログにおいて、図5に示すように各商品の商品コード欄がクリックابلエリア1～3に設定されているものとする。なお、サイト運営者は、各商品の商品コードと商品ページのURLが予め対応付けられたリストに基づいて、HTMLファイルを作成する。

[0037] サイト運営者は、ウェブカタログに対応するHTMLファイルの作成が完了すると、判定装置10において、リンクの対応付けの正誤を検査する検査モードを選択する。これにより、以下の検査処理が実行される。

[0038] まず、ページ取得部11は、ページDB4aから図4に示すHTMLファイルを取得する(S101)。

[0039] 次に、座標取得部12は、S101で取得されたHTMLファイルにおいて、mapデータの1番目のデータから、クリックابلエリア1(図5参照)の座標(x1、x2、x3、x4)を取得する(S102)。

[0040] 次に、画像取得部13は、図1に示すウェブカタログの画像データ(カタログデータ)において、S102で取得された座標(x1、x2、x3、x4)に対応するクリックابلエリア1の画像(エリア画像1)を取得する(S103)。ここでは、エリア画像1として、商品コード欄の「★1234」が取得される。

[0041] 次に、文字認識部14は、S103で取得されたエリア画像1について文字認識を行う(S104)。ここでは、エリア画像1の「★1234」が文字として認識される。

[0042] 次に、URL取得部15は、S101で取得されたHTMLファイルにおいて、S102で取得された座標(x1、x2、x3、x4)に対応するURL「http://aaa.co.jp/しょうゆ.html」を取得する(S105)。

[0043] 次に、リンク先ページ取得部16は、ウェブサーバから、S105で取得されたURL「http://aaa.co.jp/しょうゆ.html」の商品ページ(図2参照)を取得する(S106)。

[0044] 次に、文字判定部17は、S106で取得された「しょうゆ」の商品ペー

ジに、S 1 0 4 で認識された文字「★ 1 2 3 4」が含まれているか否かを判定する（S 1 0 7）。図 2 に示す例では、「しょうゆ」の商品ページに、商品コード「★ 1 2 3 4」が含まれているため、リンクの対応付けは正しいとして、S 1 0 9 に移行する。

[0045] 一方、S 1 0 6 で取得された商品ページが例えば図 8 に示す内容であった場合、ウェブカタログに記載されている商品コード「★ 1 2 3 4」と、商品ページに記載されている商品コード「★ 1 2 3 1」とが異なっているため、リンクの対応付けが誤っていると、S 1 0 8 に移行する。図 9 は、図 8 に示す商品ページに対応する HTML ファイルを示している。同図に示すように、map データの 1 番目のデータにおいて、誤った URL 「http : / / a a a . c o . j p / ドレッシング . h t m l 」が設定されている。

[0046] S 1 0 8 において報知部 1 8 がエラーメッセージを報知した後、S 1 0 9 に移行する。これによりサイト運営者は、HTML ファイルにおいてリンクの対応付けに誤りがあることを認識することができる。

[0047] S 1 0 9 では、上記検査処理を行っていないクリッカブルエリア（未検査エリア）があるか否かを判定する。具体的には、S 1 0 1 で取得された HTML ファイルにおける map データを参照して上記判定処理を行う。これにより全てのクリッカブルエリアについて上記検査処理を実行することができる。

[0048] 未検査エリアがある場合は S 1 0 2 に戻り、座標取得部 1 2 は、S 1 0 1 で取得された HTML ファイルの map データから次のクリッカブルエリアの座標を取得する。ここでは、座標取得部 1 2 は、クリッカブルエリア 2 の座標（b 1、b 2、b 3、b 4）を取得する（S 1 0 2）。以降、上記と同様の処理が行われる。未検査エリアがない場合、すなわち全てのクリッカブルエリアについて上記の処理が行われた場合は、判定装置 1 0 は検査処理を終了する。

[0049] 以上のように、本実施形態に係る判定装置 1 0 によれば、ウェブカタログの商品情報と、該商品情報のリンク先の商品ページの URL との対応付けの

正誤を容易に判定することができる。また、クリックابلエリアとURLとの対応付けが正しい場合であっても、URLで特定された商品ページ（コンテンツ）が誤っていてクリックابلエリアの情報と対応していない虞もあるが、この場合でも上記誤りを検出することができる。

[0050] [実施例2]

実施例1ではクリックابلエリアが商品コード欄に設定されているが、本発明はこれに限定されない。実施例2では、クリックابلエリアがテキスト表示欄（図5のクリックابلエリア4）に設定されている。以下では、実施例1との相違点を中心に説明する。図10は、図1に示すウェブカタログに掲載されている商品コード「★5678」の商品「Aしょうゆ」の情報と、該商品「Aしょうゆ」の商品ページを示している。

[0051] 座標取得部12は、ページ取得部11により取得されたHTMLファイル（図11参照）において、mapデータから、クリックابلエリア4の座標（s1、s2、s3、s4）（図5参照）を取得する。なお、クリックابلエリア1～3については実施例1に示した検査処理を実行してもよい。

[0052] 画像取得部13は、図10に示すウェブカタログのカタログデータ（画像データ）において、座標取得部12により取得された座標（s1、s2、s3、s4）に対応するクリックابلエリア4の画像（エリア画像4）を取得する。ここでは、エリア画像4として、テキスト表示欄が取得される。

[0053] 文字認識部14は、画像取得部13により取得されたエリア画像4について文字認識を行う。ここでは、エリア画像4のテキストが文字（単語）として認識される。具体的には、文字認識部14は、形態素解析により、例えば「しょうゆ」、「百選」、「Aしょうゆ」、「500ml」、「鰹節」、「みりん」、「まろやか」、「味わい」を認識する。

[0054] URL取得部15は、ページ取得部11により取得されたHTMLファイル（図11参照）において、座標取得部12により取得された座標（s1、s2、s3、s4）に対応するURL「http://aaa.co.jp/Aしょうゆ.html」を取得する。

- [0055] リンク先ページ取得部 16 は、ウェブサーバから、URL 取得部 15 により取得された URL 「http://aaa.co.jp/Aしょうゆ.html」の商品ページ（図 10 参照）を取得する。
- [0056] 文字判定部 17 は、リンク先ページ取得部 16 により取得された「Aしょうゆ」の商品ページに、文字認識部 14 により認識された複数の文字のうち所定割合以上の数の文字が含まれているか否かを判定する。例えば、上記の 8 単語のうち 8 割（7 単語）以上が、「Aしょうゆ」の商品ページに含まれているか否かを判定する。図 10 の例では、「Aしょうゆ」の商品ページに「百選」は含まれていないが、他の 7 単語が含まれている。よって、リンクの対応付けは正しいと判定される。
- [0057] このように、クリックابلエリアがテキスト表示欄に設定されている場合には、文字数を比較することによりリンクの対応付けの正誤を判定することができる。
- [0058] ここで、クリックابلエリアとして、商品の写真欄が追加されていてもよい。この場合は、ウェブカタログの商品の写真と、商品ページの写真とを画像認識して両者の類似度を算出し、類似度に基づいてリンクの対応付けの正誤を判断する構成とすることができる。なお、画像認識を用いた判定処理は、文字認識を用いた上記判定処理において誤りと判定された場合（S107 にて NO）に実行してもよい。
- [0059] 以上のように、判定装置 10 は、様々なクリックابلエリアについて、リンクの対応付けの正誤を判定することができる。なお、1つの商品について、商品コード欄、テキスト表示欄、写真欄の全てがクリックابلエリアに設定されていてもよい。この場合は、クリックابلエリア毎に、上述した各判定処理のうち、クリックابلエリアに対応する判定処理を行えばよい。また上記実施形態では、PDF を利用して作成したウェブカタログについて説明したが、リンクを含むウェブページ全般に本発明を適用することができる。
- [0060] [html ファイルの生成]

ページ生成部 19 の詳細について説明する。図 12 は、ページ生成部 19

の動作フロー図である。

- [0061] ページ生成部19は、ウェブカタログの元となるPDF形式の画像データに基づいて、クリックابلエリアと、これに対応するリンク先URLとを設定し、HTMLファイルを作成する処理を行う。すなわち、ページ生成部19は、クリックابلエリアの設定とリンク先URLの対応付けとを自動で行う。ここでは、図1に示すウェブカタログの画像データ（カタログデータ）を例に挙げる。
- [0062] まず、ページ生成部19は、カタログデータを取得すると（S201）、カタログデータからクリックابلエリアを特定するためのマーク（キーワード）を抽出する（S202）。例えばマークとして「★」が付されている。
- [0063] 次に、ページ生成部19は、マークに付された商品コードを、マークの座標に基づいて配置順に並べる（S203）。図13には、配置順に並べた商品コード一覧のテーブルを示している。
- [0064] 次に、ページ生成部19は、1番目の商品コードのマークの座標に基づいて、1番目のクリックابلエリア1の座標（x1、x2、x3、x4）を決定する（S204）。
- [0065] 次に、ページ生成部19は、サイト運営者により作成されたURL一覧のテーブル（図14参照）から、1番目のURL「http://aaa.co.jp/しょうゆ.html」を取得する（S205）。
- [0066] 次に、ページ生成部19は、S204で決定された1番目のクリックابلエリア1の座標（x1、x2、x3、x4）と、S205で取得された1番目のURL「http://aaa.co.jp/しょうゆ.html」とを対応付けてHTMLファイル（図15参照）に登録する（S206）。
- [0067] 続いて、S204に戻り、ページ生成部19は、2番目の商品コードのマークの座標に基づいて、2番目のクリックابلエリア2の座標（y1、y2、y3、y4）を決定する。以降、上記と同様の処理が行われ、HTMLファイルに、2番目のクリックابلエリア2の座標（y1、y2、y3、y4）と、2番目のURL「http://aaa.co.jp/リンゴ.html」

」とが対応付けられて登録される。

[0068] ページ生成部 19 は、全てのクリッカブルエリアについて上記処理を繰り返し行い、HTML ファイルに、各クリッカブルエリアの座標と各 URL とが対応付けられて登録される。これにより、図 4 に示す HTML ファイルが生成される。

[0069] ページ生成部 19 の構成によれば、クリッカブルエリアの座標と各 URL との対応付けを自動で行うことができるため、該対応付けの誤りを減らすことができる。

[0070] [商品ページの取得]

図 4 に示す HTML ファイルでは、クリッカブルエリアに対応付けられる商品ページは URL で特定されている。しかし、商品ページがウェブサーバにアップロードされる前の段階では、商品ページを URL で特定することができない。そこで、リンク先ページ取得部 16 は、URL 取得部 15 により取得された URL をローカルアドレスに変換し、変換されたアドレスに基づいて商品ページを取得する。

[0071] 具体的には、リンク先ページ取得部 16 は、複数の商品ページがローカルストレージに保存されている場合、URL を「ローカルストレージのパス＋ファイル名」に変換する。例えば、「URL : http://aaa.co.jp/しょうゆ.html」を、「C:\temp\しょうゆ.html」に変換する。なおローカルストレージとして、例えば図 16 に示すように判定装置 10 の記憶部 4 におけるリンク先ページ DB 4b とすることができる。これにより、リンク先ページ取得部 16 は、商品ページを確実に取得することができる。

[0072] 上記の実施形態では 1 ページのみのウェブカタログについて説明したが、ページ数が多いウェブカタログの場合、リンク箇所が多いため本発明を適用することでリンクの対応付け及びリンクの対応付けの誤りの判定を効率的に行うことができる。なお、上記実施形態では、本発明をウェブページに適用した場合について説明したが、ウェブページに限られない。例えば、スマー

フォンやタブレットのアプリケーションで表示され、リンクを含んだ画面（ページ）に適用してもよい。この場合、リンク先はURI（Uniform Resource Identifier）によって特定される。

符号の説明

[0073] 10 判定装置、1 通信部、2 CPU、3 メモリ、4 記憶部、4 a ページDB、4 b リンク先ページDB、11 ページ取得部、12 座標取得部、13 画像取得部、14 文字認識部、15 URL取得部、16 リンク先ページ取得部、17 文字判定部、18 報知部、19 ページ生成部。

請求の範囲

- [請求項1] U R I が対応付けられたリンクエリア内又は該リンクエリア周辺から得られるリンク元情報を取得する情報取得手段と、
- コンテンツを記憶する記憶手段から、前記リンクエリアに対応付けられた前記U R I によって特定されたコンテンツを取得するコンテンツ取得手段と、
- 前記情報取得手段により取得された前記リンク元情報と、前記コンテンツ取得手段により取得された前記コンテンツとに基づいて、前記リンクエリアと前記U R I との対応付けの正誤を判定する判定手段と、
- を含むことを特徴とする判定装置。
- [請求項2] 前記情報取得手段により取得された前記リンク元情報を文字として認識する文字認識手段をさらに含み、
- 前記判定手段は、前記文字認識手段により認識された前記文字と、前記コンテンツ取得手段により取得された前記コンテンツとに基づいて、前記リンクエリアと前記U R I との対応付けの正誤を判定することを特徴とする請求項1に記載の判定装置。
- [請求項3] 前記判定手段は、前記文字認識手段により認識された前記文字が、前記コンテンツ取得手段により取得された前記コンテンツに含まれるか否かにより、前記対応付けの正誤を判定することを特徴とする請求項2に記載の判定装置。
- [請求項4] 前記判定手段は、前記文字認識手段により認識された複数の前記文字のうち所定割合以上の数の文字が、前記コンテンツ取得手段により取得された前記コンテンツに含まれるか否かにより、前記対応付けの正誤を判定することを特徴とする請求項2に記載の判定装置。
- [請求項5] ページ内における前記リンクエリアの座標を取得する座標取得手段をさらに含み、
- 前記情報取得手段は、前記座標取得手段により取得された前記リン

クエリアの座標に基づいて、前記リンク元情報を取得することを特徴とする請求項 1 から 4 の何れか 1 項に記載の判定装置。

[請求項6] 前記対応付けが誤っている場合にエラーメッセージを報知する報知手段をさらに含んでいることを特徴とする請求項 1 から 5 の何れか 1 項に記載の判定装置。

[請求項7] ページ内の前記リンクエリアを設定するとともに、予め前記 U R I が登録されたテーブルから、設定された前記リンクエリアに対応する前記 U R I を取得し、前記リンクエリアと前記 U R I とを対応付ける生成手段をさらに含んでいることを特徴とする請求項 1 から 6 の何れか 1 項に記載の判定装置。

[請求項8] 前記テーブルにおいて、前記 U R I は、前記リンクエリアの配置順に対応して登録されていることを特徴とする請求項 7 に記載の判定装置。

[請求項9] U R I が対応付けられたリンクエリア内又は該リンクエリア周辺から得られるリンク元情報を取得する情報取得ステップと、

コンテンツを記憶する記憶手段から、前記リンクエリアに対応付けられた前記 U R I によって特定されたコンテンツを取得するコンテンツ取得ステップと、

前記情報取得ステップにより取得された前記リンク元情報と、前記コンテンツ取得ステップにより取得された前記コンテンツとに基づいて、前記リンクエリアと前記 U R I との対応付けの正誤を判定する判定ステップと、

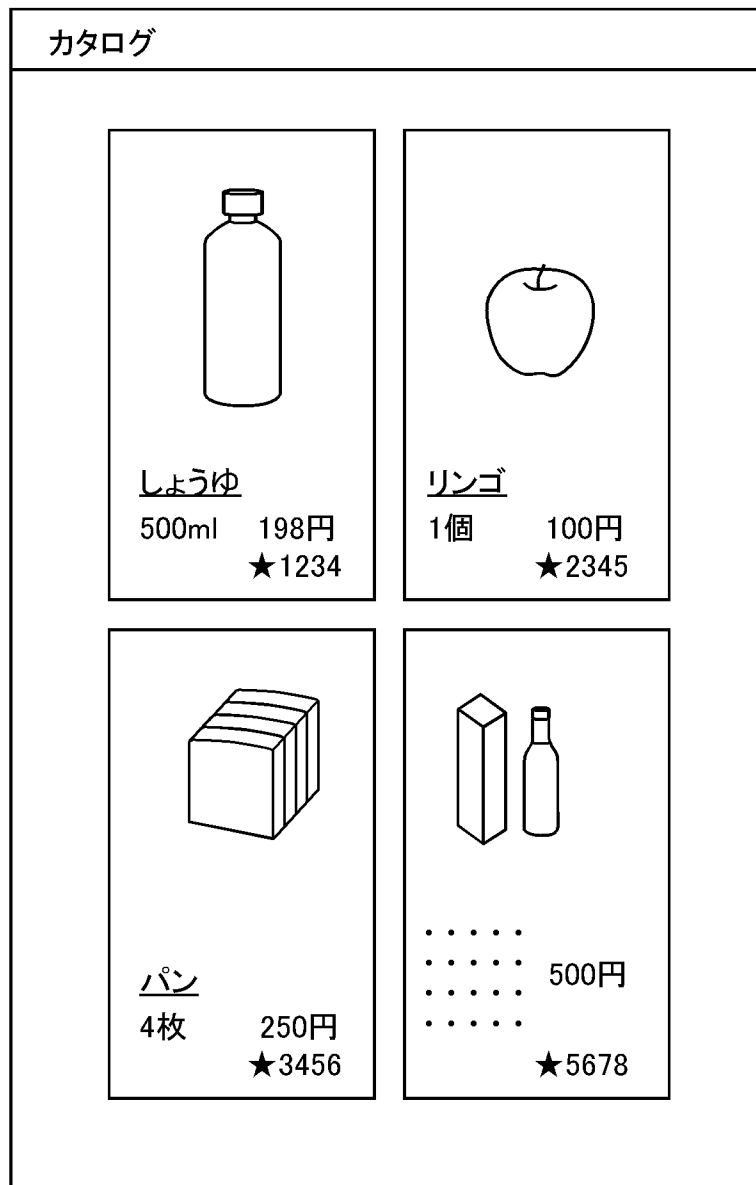
を含むことを特徴とする判定方法。

[請求項10] U R I が対応付けられたリンクエリア内又は該リンクエリア周辺から得られるリンク元情報を取得する情報取得手段、

コンテンツを記憶する記憶手段から、前記リンクエリアに対応付けられた前記 U R I によって特定されたコンテンツを取得するコンテンツ取得手段、及び、

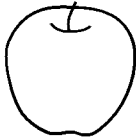
前記情報取得手段により取得された前記リンク元情報と、前記コンテンツ取得手段により取得された前記コンテンツとに基づいて、前記リンクエリアと前記ＵＲＬとの対応付けの正誤を判定する判定手段、
としてコンピュータを機能させるためのプログラム。

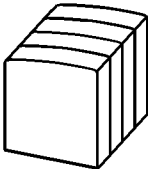
[図1]



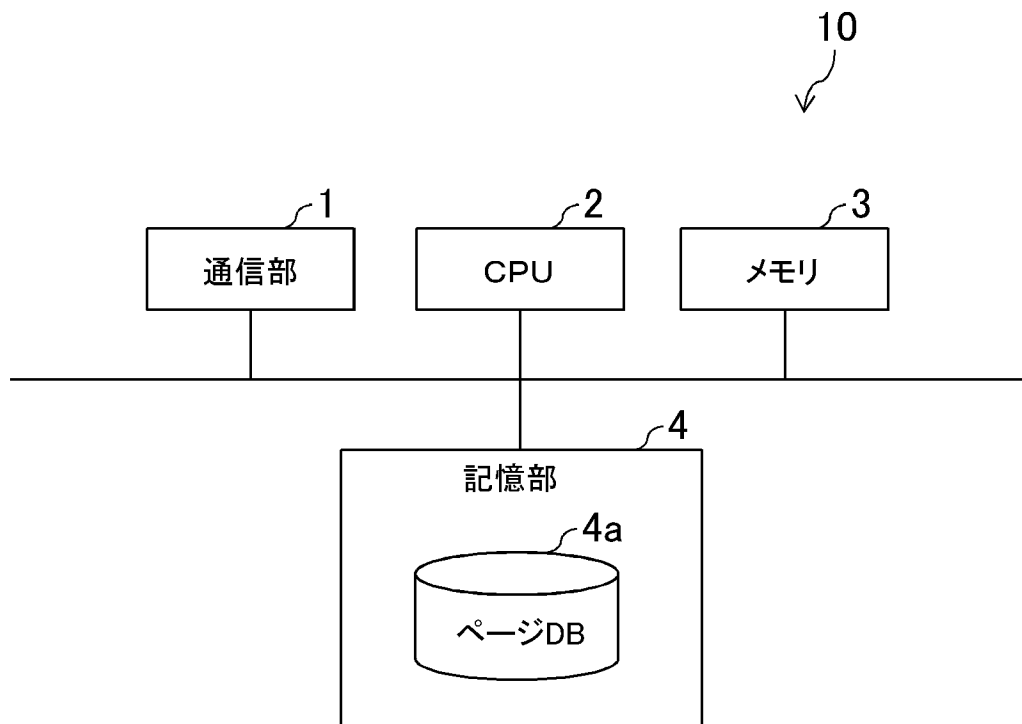
[図2]

しょうゆ	
	しょうゆ 500ml 198円 メーカー名: abc 原材料: 大豆、小麦
★1234	購入

リンゴ	
	リンゴ 1個 100円 メーカー名: bcd 原産地: 青森県
★2345	購入

パン	
	パン 4枚 250円 メーカー名: cde 原材料: 小麦粉、砂糖、 塩、イースト
★3456	購入

[図3]



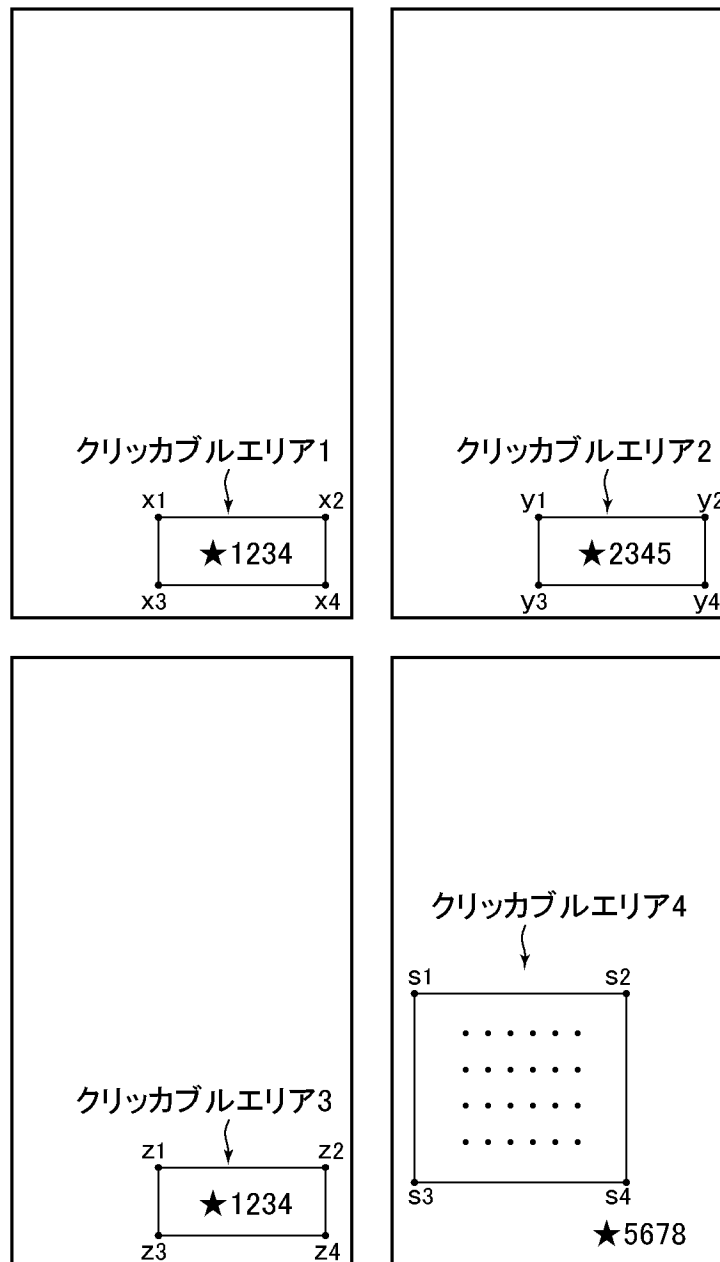
[図4]

```
<html>
<h1>カタログ</h1>

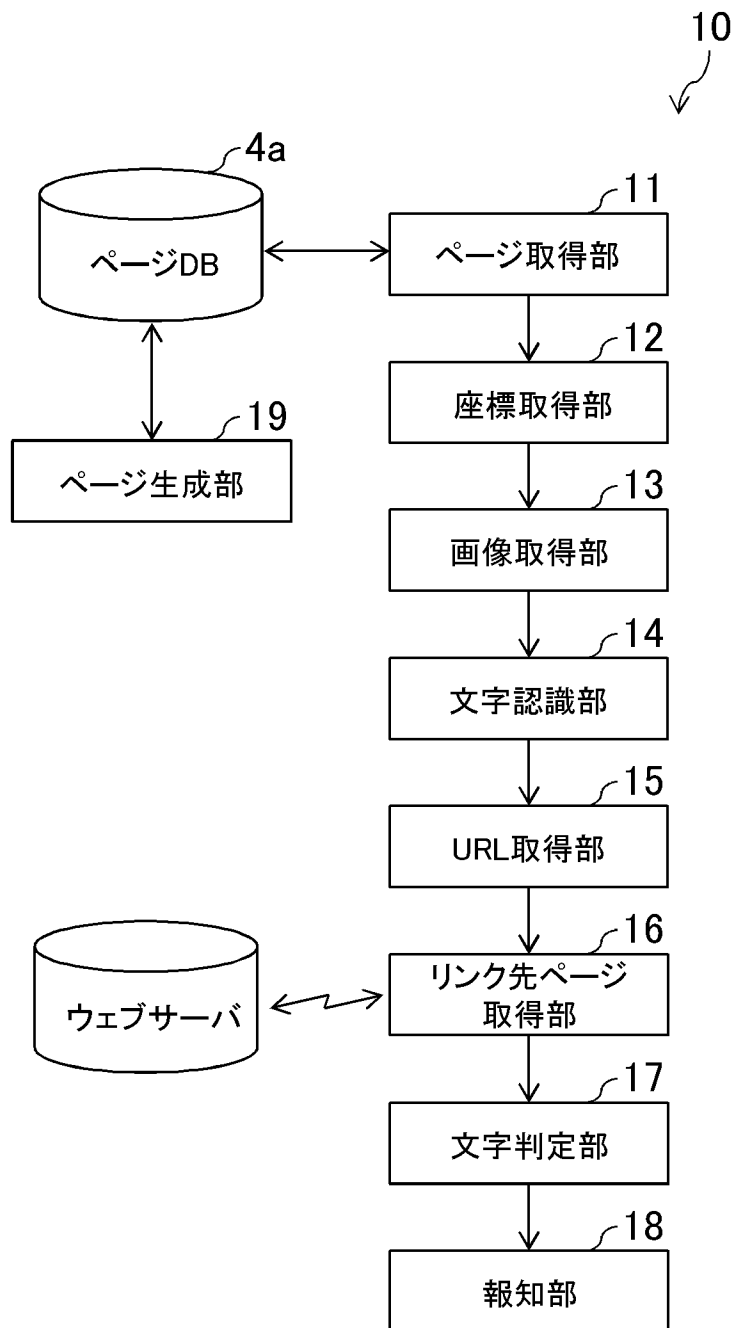
<map name="rtmap">
  <area href="しょうゆ.html" shape="rect" coords="x1,x2,x3,x4">
  <area href="リンゴ.html" shape="rect" coords="y1,y2,y3,y4">
  <area href="パン.html" shape="rect" coords="z1,z2,z3,z4">
</map>
</html>
```

mapデータ

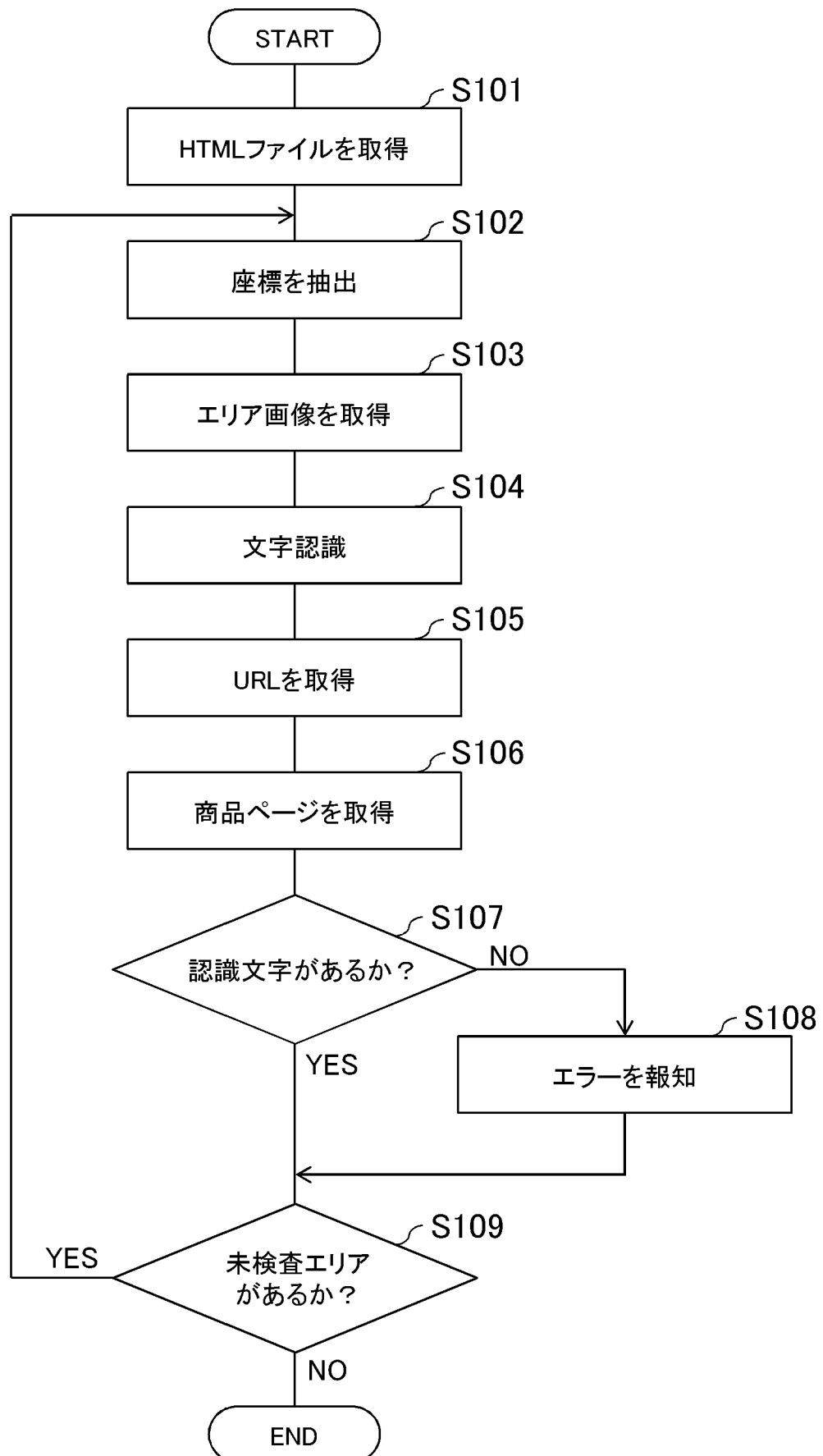
[図5]



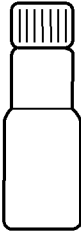
[図6]



[図7]



[図8]

ドレッシング	
	ドレッシング 200ml 500円 メーカー名 : def 原材料 : 玉ねぎ、にんにく、食塩、こしょう
★1231	購入

[図9]

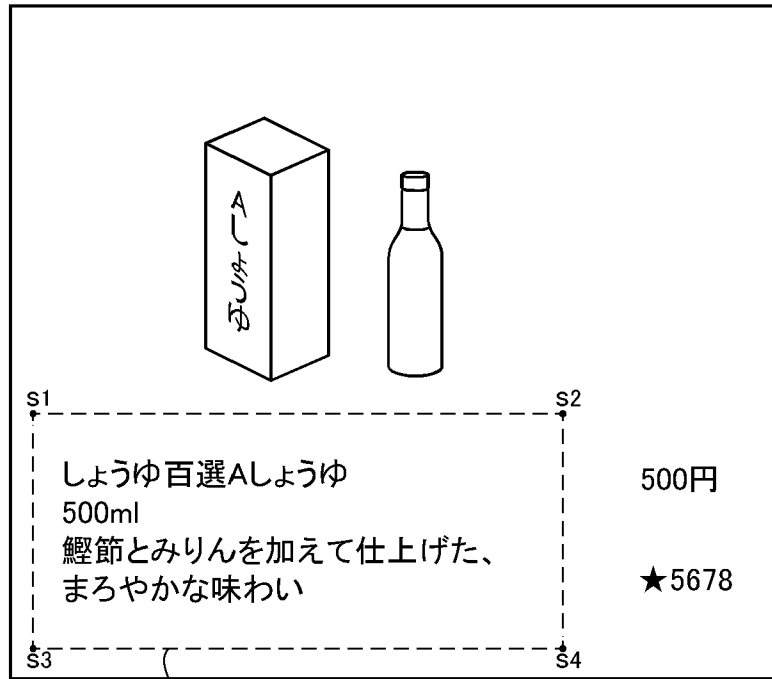
```
<html>
<h1>カタログ</h1>

<map name="rtmap">
  <area href="ドレッシング.html" shape="rect" coords="x1,x2,x3,x4">
  <area href="リンゴ.html" shape="rect" coords="y1,y2,y3,y4">
  <area href="パン.html" shape="rect" coords="z1,z2,z3,z4">
</map>
</html>
```

mapデータ

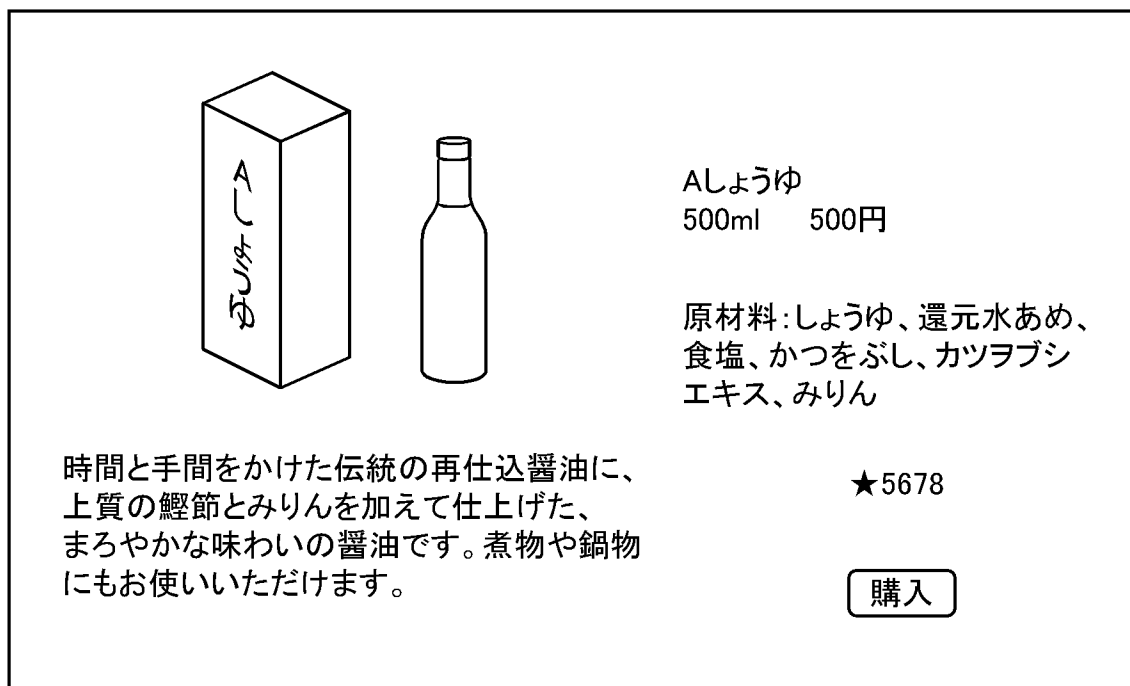
[図10]

Aしょうゆの情報



クリックブルエリア4

Aしょうゆのページ



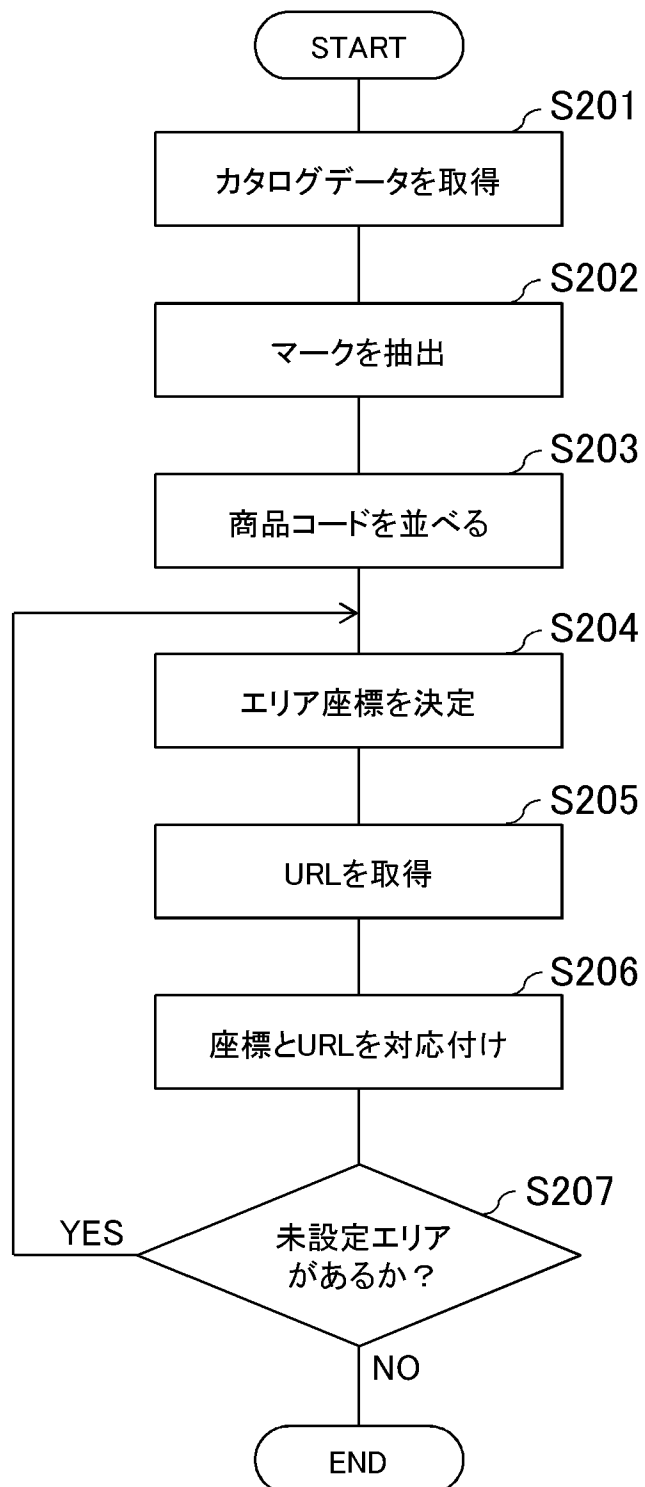
[図11]

```
<html>
<h1>カタログ</h1>

<map name="rtmap">
  <area href="しよっぴ.html" shape="rect" coords="x1,x2,x3,x4">
  <area href="リンゴ.html" shape="rect" coords="y1,y2,y3,y4">
  <area href="パン.html" shape="rect" coords="z1,z2,z3,z4">
  <area href="Alしょうひ.html" shape="rect" coords="s1,s2,s3,s4">
</map>
</html>
```

mapデータ

[図12]



[図13]

順番	商品コード
1	1234
2	2345
3	3456
⋮	⋮

[図14]

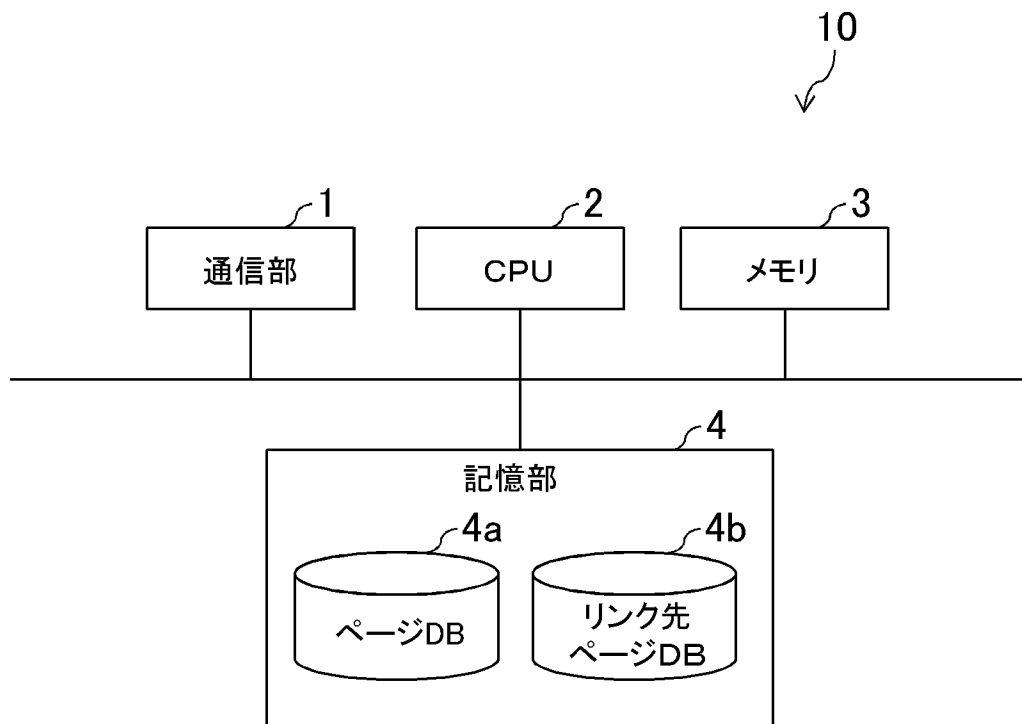
順番	URL
1	しょうゆ. html
2	リンゴ. html
3	パン. html
⋮	⋮

[図15]

登録

```
<html>  
<h1>カタログ</h1>  
  
<map name="rtmap">  
→ <area href="しょうゆ.html" shape="rect" coords="x1,x2,x3,x4">  
</map>  
</html>
```

[図16]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/067942

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/21(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F17/21, G06F12/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2007-188356 A (International Business Machines Corp.), 26 July 2007 (26.07.2007), entire text; all drawings & US 2008/0172220 A1 & CN 101000628 A	1-10
Y	JP 2004-139304 A (NEC Corp.), 13 May 2004 (13.05.2004), paragraph [0076] (Family: none)	1-10
Y	JP 9-44499 A (Matsushita Electric Industrial Co., Ltd.), 14 February 1997 (14.02.1997), paragraphs [0012], [0018] to [0019]; fig. 5 (Family: none)	7-8

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
23 July, 2013 (23.07.13)

Date of mailing of the international search report
30 July, 2013 (30.07.13)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (I P C))

Int.Cl. G06F17/21 (2006.01) i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (I P C))

Int.Cl. G06F17/21、G06F12/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 3 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 3 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 3 年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2007-188356 A (インターナショナル・ビジネス・マシーンズ・コーポレーション) 2007.07.26, 全文全図 & US 2008/0172220 A1 & CN 101000628 A	1-10
Y	JP 2004-139304 A (日本電気株式会社) 2004.05.13, 段落【0076】 (ファミリーなし)	1-10
Y	JP 9-44499 A (松下電器産業株式会社) 1997.02.14, 段落【0012】、【0018】 - 【0019】、図5 (ファミリーなし)	7-8

C 欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
「O」 口頭による開示、使用、展示等に言及する文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」 同一パテントファミリー文献

国際調査を完了した日

2 3 . 0 7 . 2 0 1 3

国際調査報告の発送日

3 0 . 0 7 . 2 0 1 3

国際調査機関の名称及びあて先

日本国特許庁 (I S A / J P)
郵便番号 1 0 0 - 8 9 1 5
東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

長 由紀子

5 M

4 2 3 3

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 9 9