(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2012 (18.10.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/141934 A1

(51) International Patent Classification:
G06F 7/00 (2006.01)

(21) International Application Number:
PCT/US2012/031875

(22) International Filing Date:
2 April 2012 (02.04.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201110093430.9 14 April 2011 (14.04.2011) CN

(71) Applicant *(for all designated States except US)*: **ALIBABA GROUP HOLDING LIMITED** [—/US]; Fourth Floor, One Capital Place, P.O. Box 847, Grand Cayman (KY).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MIAO, Wen, Yan** [CN/CN]; C/o Alibaba Group Legal Department, 10/f, Building A, The West, Lake International Plaza Of S&t., No. 391 Wen'er Road, Hangzhou, 310013 (CN). **QIAN, Jian, Ping** [CN/CN]; C/o Alibaba Group Legal Department, 10/f, Building A, The West, Lake International Plaza Of S&t, No. 391 Wen'er Road, Hangzhou, 310013 (CN).

(74) Agent: **LEE, Lewis, C.**; Lee & Hayes, PLLC, 601 W. Riverside Ave, Suite 1400, Spokane, WA 99201 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, KM, 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, 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) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: METHOD AND APPARATUS OF DETERMINING PRODUCT CATEGORY INFORMATION

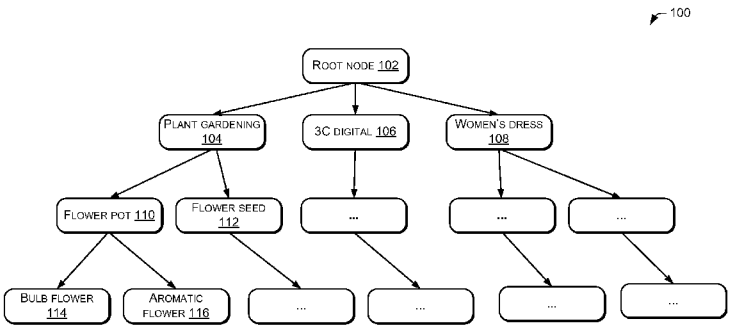


FIG. 1

(57) **Abstract**: A user may submit product title information to a server. The server may generate a phrase based on the product title information. The server may then search a database to find relevancies between the phrase and product categories corresponding to multiple nodes in a product category tree. Based on the relevancies, the server may select a node from the multiple nodes. The server may associate the product title information with the node corresponding to a product category when the node is a leaf node of the product category tree.

WO 2012/141934 A1

METHOD AND APPARATUS OF DETERMINING PRODUCT CATEGORY INFORMATION

CROSS REFERENCE TO RELATED PATENT APPLICATIONS

This application claims priority to Chinese Patent Application No. 201110093430.9, filed on April 14, 2011, entitled "METHOD AND APPARATUS OF DETERMINING PRODUCT CATEGORY INFORMATION," which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

This disclosure relates to the field of data processing. More specifically, the disclosure relates to a method and an apparatus of determining product category information.

BACKGROUND

When a user uploads information of products through a server, the user needs to determine a category of the uploaded product according to product category information provided by the server, and associate the category with the information of the products. However, conventional technologies of determining product category information may present some problems (e.g., inaccurate results and inefficiency). For example, it may be difficult to select proper category when the product category information provided by the server is complex.

SUMMARY OF THE DISCLOSURE

This disclosure provides methods and devices for determining product category

information. In some embodiments, a user may submit product title information to a server. The server may generate a phrase based on the product title information. The server may then search a database to find relevancies between the phrase and product categories corresponding to multiple nodes in a product category tree. Based on the relevancies, the server may select a node from the multiple nodes. The server may associate the product title information with the node corresponding to a product category when the node is a leaf node of the product category tree.

In some embodiments, the server may select a node of a product category tree. The node may correspond to a product category that includes multiple product titles. The server may generate a phrase based on a product title of the multiple product titles, and then calculate a relevancy between the phrase and the product category based on relationship between the phrase and the multiple product titles.

BRIEF DESCRIPTION OF THE DRAWINGS

The Detailed Description is described with reference to the accompanying figures. The use of the same reference numbers in different figures indicates similar or identical items.

FIG. 1 is a block diagram of an illustrative structure of an example product category tree.

FIG. 2 is a block diagram of an illustrative architecture to determine product category information.

FIG. 3 is a flow diagram of an illustrative process to determine product category information.

FIG. 4 is a flow diagram of an illustrative process to calculate relevancy between

product information and product category.

FIG. 5 is a flow diagram of another illustrative process to determine product category information.

FIG. 6 is a block diagram of an illustrative computing device that may be deployed in the environment shown in FIG. 2.

DETAILED DESCRIPTION

The disclosure introduces a method, a system and an apparatus of determining product category information for improving accuracy and efficiency of searching product categories with which a product may be associated. A relevancy between product information and corresponding product category information may be calculated and stored in a database. The product information may be provided by a user, and the corresponding product category information may be provided by a service provider. As one example, the relevancy may be stored in the database as Product_ID-Product_Title_Information-Product_Category_ID.

The product category may be a product category corresponding to a leaf node of a product category tree. In the product category tree, there are higher level nodes of the leaf node (e.g., parent nodes), while there are no lower level nodes of the leaf node (e.g., child nodes). The product information may be associated with the product category corresponding to the leaf node. In addition, because of the relationship between parent nodes and their child nodes, each corresponding node is determined from a root node to child nodes connected to the root node, and eventually to a corresponding leaf node during the determination of the product category. Therefore, in a product category tree, the product information may be

associated with the leaf node, and product categories corresponding to parent nodes of the leaf node may include the product information.

FIG. 1 is a block diagram of an illustrative structure of an example product category tree 100. The product category tree 100 may include a root node 102, which may include three child nodes: a plant gardening node 104, 3C digital node 106, and women's dress node 108, which correspond to three product categories. The plant gardening node 104 may further include two child nodes: a flower pot node 110 and a flower seed node 112. The flower pot node 110 may further include two child nodes: a bulb flower node 114 and an aromatic flower node 116.

The bulb flower node 114 and the aromatic flower node 116 may be leaf nodes, which correspond to product categories of bulb flowers and aromatic flowers respectively. Product information may be associated with product categories corresponding to the bulb flower node 114 and the aromatic flower node 116. After the product information is associated with a leaf node (e.g., the bulb flower node 114), associated higher-level nodes of this leaf node (e.g., the flower pot node 110 and plant gardening node 104) may include the product information as well. For example, a path from the leaf node to a particular associated higher-level node in the product category tree is reserved in the database after the product information is associated with the leaf node. Therefore, a relevancy between the product information and associated product category information reserved in the database can be shown as: Product_ID- Product_Title_Information-Associated_Root_Node_ID-Associated_Parent_Node_ID.

FIG. 2 is a block diagram of an illustrative architecture 200 to determine product category information. Under the architecture 200, a user 202 may, via a user

device 204, log in and provide product information to a server 206, which may provide a product category tree to the user 202. The user 202 may choose a product category corresponding to a leaf node by selecting children or grandchildren nodes of the root node, and then the server 206 may associate the product information to the product category. The server 206 can be implemented as a web server. After the server 206 associate the product information to the product category, the server 206 may transmit related information to a database 208 for storage. In some embodiments, a cloud computing platform 210 may be implemented. For example, the cloud computing platform 210 may include Hadoop.

The cloud computing platform 210 may include a distributed analysis system 212 and a real-time analysis system 214. In some embodiments, the cloud computing platform 210 may receive periodically the related information from the database 208. The cloud computing platform 210 may store the product information and the associated product category information, analyze the information, and determine the relevancy between the product information and the product category. For example, the cloud computing platform 210 may determine a relevancy between each phase of the product information and each corresponding product category of the product category tree. In some embodiments, the distributed analysis system 212 and the real-time analysis system 214 may be implemented to determine the relevancy. The distributed analysis system 212 may be an off-line distributed analysis system, and the real-time analysis system 214 may be an online real-time analysis system. The cloud computing platform 210 may further provide information associated with relevancies to the server 206 to determine the product category information.

FIG. 3 is a flow diagram of an illustrative process 300 to determine product category information. At 302, the server 206 may receive product information, which may be submitted by the user 202. The product information may include product title information and other description information of products. In some embodiments, the server 206 may receive the product information from other servers while incorporating the product information in other servers into the database.

At 304, the server 206 may conduct word segmentation on the product title information to determine each phrase including one or more word segments and corresponding to the product title. In some embodiments, the word segment may be a word and/or character, or multiple words and/or characters. At 306, the server 206 may search a relevant value between each phrase and each child node (e.g., the plant gardening node 104, the 3C digital node 106 and the woman's dress node 108) of a root node (e.g., the root node 102). The server 206 may determine relevancies between phrases and product categories based on relevant values between phrases and each child node. Therefore, the server 206 may determine a relevancy between each phrase and each product category.

At 308, the server 206 may determine a particular node corresponding to the product information based on the relevancy between each phrase and each child node. In some embodiments, the server 206 may select child nodes that have relevant values regarding a certain phrase grater than a predetermined relevant value. In some embodiments, the server 206 may further determine a sum relevancy between all phrases derived from the product title information and a product category corresponding to each selected child node. In some embodiments, the

server 206 may select a product category corresponding to the product title information such that the product category has the greatest sum relevancy than other product category.

At 310, the server 206 may set a child node of the root node that corresponds to the product category as a parent node. The server 206 may then associate products corresponding to the product title information to a product category corresponding to a leaf node derived from the parent node.

FIG. 4 is a flow diagram of an illustrative process 400 to calculate relevancy between product information and product category. At 402, the server 206 may determine a parent node corresponding to a node of each product category. In some embodiments, a product category may be the category to which each product belongs. Each product category may correspond to a node of a product category tree, and the node with which the product information associated with may be the leaf node. In the product category tree 100, starting from the root node of the product category tree, the nodes are connected with each other. The root node has child nodes, leaf nodes only have parent nodes, and middle nodes have parent nodes and child nodes.

At 404, the server 206 may determine phrases corresponding to product title information of each produce included in the node. In some embodiments, the server 206 may conduct word segmentation on the product title information to determine one or more phrases corresponding to the product title. Each of the phrases may include one or more word segments. For each phrase, the server 206 may determine a relevant value between the phrase and the parent node based on a number that the phrase is contained in product titles included in the product category

corresponding to the parent node. The relevant value may determined based further on a number that the phrase is contained in product titles included in the product category corresponding to the node, and numbers of products included in product categories corresponding to the node and the parent node respectively.

For example, the server 206 may determine a weighting of the phrase and the corresponding node based on the number that the phrase is contained in product titles included in the product category corresponding to the node, and a number of products included in a product category corresponding to the node. The server 206 may determine another weighting of the phrase and other product categories corresponding to the parent node based on numbers of products included in product categories corresponding to the node and the parent node respectively. Based on the two weightings, the server 206 may calculate the relevant value between the phrase and the corresponding product category. For each node, the relevant value between each phrase of the product title information and the corresponding product category may be stored.

For example, to determine a relevancy between each phrase and the flower pot 110, the server 206 may determine the plant gardening 104 as the parent node. For product title information included in the flower pot 110, the server 206 may calculate the relevant value between each phrase of the product title information and the flower pot 110. To determine a relevancy between each phrase and the plant gardening 104, the server 206 may identify the parent node of the plant gardening 104 as the root node 102. For product title information of each product included in the plant gardening node 104, the server 206 may calculate the relevant value between each phrase and the plant gardening 104. In some embodiments, the

server 206 may calculate relevant values starting from leaf nodes to the root node or from the root node to the leaf nodes.

Suppose that a product with an ID "1000" is associated with the bulb flower node 114. As results, according to the product category tree 100, the bulb flower node 114 is the leaf node, and the flower pot node 110 is its parent node. Further suppose that the title information of the product 1000 is "Potable Hydroponics Purple Hyacinth Bulbs." As results, the server 206 may conduct word segmentation on the title information, generate phrases based on "Hyacinth," and determine a relevant value between phrases including "Hyacinth" and the bulb flower node 114.

Further suppose that the number of products related to the product category corresponding to the flower pot node 110 is 1,623,912 and the number of products related to the product category corresponding to the bulb flower node 114 is 104,286. In addition, the number of occurrences of the phrase "hyacinth" in the product title information of products of the bulb flower node 114 is 6,588, and the number in the product title information of products of the flower pot node 110 is 20,683. As results, a weighting of the product category of the bulb flower node 114 regarding the phrase "hyacinth" is 6.3% (i.e., 6588/104286). Another weighting of other products of the flower pot node 110 is 0.1% (i.e., (20683-6588)/(1623912-104286)). Therefore the relevant value between the phrase "hyacinth" and the product category of the bulb flower 114 is 85.3%. (i.e., [6588/104286-(20683-6588)/(1623912-104286)]/(6588/104286)). Similarly, the server 206 may determine the relevant value between the phrase "hyacinth" and the product category of the flower pot node 110.

In some embodiments, the phrase may include multiple word segments (e.g.,

"Purple Hyacinth" or "Hyacinth Hyacinth Bulbs"). For example, for the digital product and the digital product accessory of 3C electronics, relevancies between different words of product title information and different product categories may be similar. It may be difficult to determine to which product category a product belongs based on a single word segment for each product category. For example, for the product title information including "genuine Samsung Galaxy S i9000 mobile phone battery", it can determine that a relevancy between the product category "mobile phone" and the product information, and another relevancy between the product category "mobile phone battery" and the product information, are substantially similar. To solve the problem, the server 206 may determine a relevancy between a phrase comprising at least two word segments for each product category. For example, the server 206 may determine a relevancy between the phrase "mobile phone battery" and the product category "mobile phone battery."

In some embodiments, a relevancy between a phrase comprising a word segment and each product category may be set as a one-order vector, and a relevancy between a phrase comprising at least two words and each product category may be set as a two-order vector.

After determining and storing the relevant value between each phrase and each product category, the server 206 may determine a product category corresponding to product title information that is provided by, for example, the user 202.

FIG. 5 is a flow diagram of another illustrative process 500 to determine product category information. At 502, the server 206 may receive product information that may include product title information. For example, the user device

204 may transmit the product information to the server 206, and request for product category determination.

At 504, the server 206 may determine word segments based on the received product title information. Each word segment may be used for product category determination. In some embodiments, the server 206 may unify the product title information and determine a character standard corresponding to each character of the product title information. For example, when the product information has both upper case letters and lower case letters, the server 206 may convert lower case letters upper case letters if the standard is upper case. The server may then conduct word segmentation on the product title information, and remove certain word segment based on a non-useable word segment table to obtain word segments for product category determination.

At 506, the server 206 may determine a segment relevancy between a word segment and each child node of the root node or a parent node. The server 206 may determine the segment relevancy based on stored relevant values between phrases and each child node.

At 508, the server 206 may determine whether the segment relevancy falls below a predetermined threshold. If the segment relevancy is less than the threshold (i.e., the "YES" branch from 508), the server 206 may determine whether the word segment is the last word of the product title information at 509. If the server 206 determine that the word segment is the last word segment of the product title information (i.e., the "YES" branch from 510), the operations 504 to 510 may be performed by a loop process (via dashed line from the operation 510 that leads back to the operation 504). The server 206 may select another word segment for product

category determination. If the word segment is not the last word segment (i.e., the "NO" branch from 510), the server 206 may generate a phrase including the word segment and at least another word segment after the word segment at 512.

If the segment relevancy is not less than the threshold (i.e., the "NO" branch from 508), the server 206 may select child nodes corresponding to each word segment at 514. The server may select child nodes that have relevant values greater than a predetermine value regarding the word segment. In some embodiments, the server may select child nodes having relevant values greater than other child nodes regarding the word segment. In some embodiments, the server 206 may search a relevancy regarding the phrase generated from the operation 512, and then determine the child node.

At 516, the server 206 may calculate one or more sum relevancies for a product category corresponding to each selected child node and the product title information. In some embodiments, the server 206 may determine a first sum relevancy based on relevancies between each word segment of the product title information and the selected product category. In some embodiments, the server 206 may determine a second sum relevancy based on relevancies between each phrase derived from the word segments and the selected product category.

At 518, the server 206 may determine a particular product category corresponding to the product title information based on sum relevancies. This particular product category may have the greatest sum relevancy. In some embodiments, the server 206 may calculate a total sum relevancy based on the first sum relevancy and the second sum relevancy and the particular product category may have the greatest total sum relevancy.

At 520, the server 206 may set the node corresponding to the particular product category as a parent node and then determine corresponding child nodes. The operations 506 to 520 may be performed by a loop process (via dashed line from the operation 520 that leads back to the operation 506) until the product title information is associated to a product category corresponding to a leaf node.

For example, suppose that product title information of a product comprises word segments A, B, and C, and the product category setting number is 3. In addition, the relevance between each word segment and each product category contained in the product title information is provided in table 1 below.

	R1	R2	R3	R4	R5
A	50%	83%	62%		
B	40%		20%		37%
C	42%	57%		32%	
AB	20%	40%			30%
AC	10%	30%	40%		

Further suppose that the product categories corresponding to child nodes contained in the root node are R1, R2, R3, the relevance between the word A and product categories R1, R2 and R3 are determined as 50%, 83% and 62%, the relevance between the word B and product categories R1, R3 and R5 are determined as 40%, 20% and 37%, and the relevance between the word C and product categories R1, R2 and R4 are determined as 42%, 57% and 32%. Suppose that the set relevance threshold is 70%.

Accordingly, the relevancy between the phrase constituted by word A and word B and the product categories R1, R2 and R5 are determined as 20%, 40% and 30%, and the relevancy between the phrase constituted by word A and word C and

the product categories R1, R2 and R3 are determined as 10%, 30% and 40%. Furthermore, the first sum relevancy between the product title information and the product category R1 is 132%, the first sum relevancy between the product title information and the product category R2 is 140%, the first sum relevancy between the product title information and the product category R3 is 82%, the first sum relevancy between the product title information and the product category R4 is 32%, and the first sum relevancy between the product title information and the product category R5 is 37%.

In addition, the second sum relevancy between the product title information and the product category R1 is 30%, the second sum relevancy between the product title information and the product category R2 is 70%, the second sum relevancy between the product title information and the product category R3 is 40%, and the second sum relevancy between the product title information and the product category R5 is 30%.

Accordingly, the sum relevancy between the product title information and the product category R1 is 162%; the sum relevancy between the product title information and the product category R2 is 210%; the sum relevancy between the product title information and the product category R3 is 122%; the sum relevancy between the product title information and the product category R4 is 32%; and the sum relevancy between the product title information and the product category R5 is 67%. Therefore, the node comprising the product information is the node corresponding to the product category R2. The server 206 may then set the node as a parent node, determine child nodes of the parent node, and determines another node of the newly determined child nodes corresponding to the product title

information until the product title information is associated to a product category corresponding to a leaf node.

FIG. 6 is a block diagram of an illustrative computing device 600 that may be deployed in the environment shown in FIG. 2. The server 206 may be configured as any suitable server(s). In one exemplary configuration, the server 206 include one or more processors 602, input/output interfaces 604, network interface 606, and memory 608.

The memory 608 may include computer-readable media in the form of volatile memory, such as random-access memory (RAM) and/or non-volatile memory, such as read only memory (ROM) or flash RAM. The memory 608 is an example of computer-readable media.

Computer-readable media includes volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules, or other data. Examples of computer storage media include, but are not limited to, phase change memory (PRAM), static random-access memory (SRAM), dynamic random-access memory (DRAM), other types of random-access memory (RAM), read-only memory (ROM), electrically erasable programmable read-only memory (EEPROM), flash memory or other memory technology, compact disk read-only memory (CD-ROM), digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other non-transmission medium that can be used to store information for access by a computing device. As defined herein, computer-readable media does not include transitory media such as modulated data signals and carrier

waves.

Turning to the memory 608 in more detail, the memory 608 may store a mode determination module 610, a control module 612, a relevancy determination module 614, a segmentation module 616, a searching module 618, and a product category determination module 620. The node determination module 610 may determine a node (e.g., child nodes and parent nodes) corresponding to a product within a product category tree.

The control module 612 may conduct word segmentation on the product title information and determine each phrase corresponding to product title information. The phrase may include at least a word segment. For each phrase, the control module 612 may determine a relevancy between the phrase and a product category corresponding to the node according to an occurrence of the phrase in product title information contained in a product category corresponding to the corresponding parent node, an occurrence of the phrase in product title information contained in a product category corresponding to the node, a number of the product contained in a product category corresponding to the node and a number of the product contained in a product category corresponding to the parent node.

In some embodiments, the server 206 may store relevance tables. The tables may include a table having relevancies between each word and each product category, and another table having relevancies between each phrase constituted by at least two words and each product category. Each phrase may include one or more word segments. The server 206 may determine the relevance between each phrase and each child node according to each child node connected to the root node and the preserved relevance between each child node and each phrase. The server 206

may determine the child node including the product title information according to the relevance between each phrase of the product title information and each child node. After that, the server 206 may set the child node as a parent node and determine child nodes of the parent node. Accordingly, the relevancy between each phrase of the product title information and each child node may be determined and the child node containing the product title information may be determined until the product title information of a product is associated with to a product category corresponding to a leaf node.

In some embodiments, the control module 612 may determine a first weighting between the phrase and a product category corresponding to the node according to an occurrence of the phrase in product title information contained in a product category corresponding to the node and a product number contained in a product category corresponding to the node. The control module 612 may determine a second weighting between the phrase and a product category corresponding to the father node according to an occurrence of the phrase in product title information contained in a product category corresponding to the father node, an occurrence of the phrase in product title information contained in a product category corresponding to the node, a number of the product contained in a product category corresponding to the node and a number of the product contained in a product category corresponding to the parent node. The control module 612 may determine a difference between the first weighting and the second weighting, and determine the relevancy between the phrase and a product category corresponding to the node according to a quotient of the difference and the first weighting.

In some embodiments, the control module 612 may determine the first

weight according to a quotient of an occurrence of the phrase in product title information contained in a product category corresponding to the node and a product number contained in a product category corresponding to the node.

In some embodiments, the control module 612 may determine a first difference between an occurrence of the phrase in product title information contained in a product category corresponding to the father node and an occurrence of the phrase in product title information contained in a product category corresponding to the node. The control module 612 may also determine a second difference between a product number contained in a product category corresponding to the father node and a product number contained in a product category corresponding to the node, and then determine the second weight according to a quotient of the first difference and the second difference.

In some embodiments, a relevancy between a phrase and a product category may be stored in the server 206. The server 206 may search and determine the product category to which the product belongs by the on-line real-time prediction system. The on-line on-time prediction system may be included in the server 206. For example, the server 206 may include an Apache module. The Apache module may be implemented by the control module 612 and the node determination 610 as well as the relevancy determination module 614.

The relevancy determination module 614 may store relevancies between phrases and product categories corresponding to nodes of the product category tree. The segmentation module 616 may receive product title information, conduct word segmentation on the product title information and determine each phrase corresponding to the product title information.

The searching module 618 may search relevancies between phrases and child nodes according to relevant values between reserved phrases and product categories corresponding to nodes. The product category determination module may determine a node containing the product information according to searched relevancies between phrases and child nodes, and set a child node as a parent node.

In some embodiments, the searching module 618 may determine a child node corresponding to each product category contained in the phrase according to the relevancy between each phrase reserved and each product category. The search module 618 may set the relevance between the stored phrase and a product category corresponding to the child node as a relevancy between the searched phrase and the child node. In some embodiments, the searching module 618 may determine whether the searched relevance between the phrase and the child node is greater than a preset relevance threshold.

The product category determination module 620 may determine a child node of the parent node until the product title information of a product is associated with a product category corresponding to a leaf node. In some embodiments, the product category determination module 620 may select child nodes having relevancies greater than a predetermined value corresponding to the searched phrase. The product category determination module 620 may determine a sum relevancy between the product title information and the product category based on a relevancy between each phrase of the product title information and the product category. The product category determination module 620 may determine a product category having the greatest sum relevance as a product category corresponding to a node containing the product information.

The embodiments in this disclosure are merely for illustrating purposes and are not intended to limit the scope of this disclosure. A person having ordinary skill in the art would be able to make changes and alterations to embodiments provided in this disclosure. Any changes and alterations that persons with ordinary skill in the art would appreciate fall within the scope of this disclosure.

CLAIMS

What is claimed is:

1. One or more computer-readable media storing computer-executable instructions that, when executed by one or more processors, performs acts comprising:

receiving product information associated with a product;

conducting word segmentation on the product information to generate a phrase that includes one or more word segments;

finding relevancies between the phrase and product categories corresponding to multiple nodes having one parent node in a product category tree, each of the multiple nodes corresponding to one product category; and

selecting a node from the multiple nodes based on the relevancies.

2. The one or more computer-readable media of claim 1, wherein the one parent node is a root node of the product category tree.

3. The one or more computer-readable media of claim 1, wherein the acts further comprise:

determining whether the node is a leaf node of the product category tree;

in the event that the node is the leaf node:

associating the product information with a product category corresponding to the node, the product information including title information of the product;

in the event that the node is not the leaf node:

determining that the node has one or more leaf nodes;

determining a leaf node of the one or more leaf nodes based on relevancies between the phrase and product categories corresponding to the one or more leaf nodes; and

associating the title information with a product category corresponding to the leaf node.

4. The one or more computer-readable media of claim 3, wherein the acts further comprise:

in the event that the node is not the leaf node:

determining that the node does not have a leaf node; and

determining one or more child nodes of the node.

5. The one or more computer-readable media of claim 1, wherein the acts further comprise:

determining whether the relevancies are less than a predetermined threshold;

in the event that the relevancies are less than the predetermined threshold, generating another phrase based on the one or more segments; and

finding other relevancies between the other phrase and the product categories corresponding to the multiple nodes.

6. The one or more computer-readable media of claim 1, wherein the selecting the node comprises:

selecting one or more nodes from the multiple nodes, the one or more nodes

corresponding to product categories that have relevancies with the phrase greater than a predetermined value;

calculating one or more sum relevancies between the phrase and the product categories corresponding to the one or more nodes; and

selecting the node from the one or more nodes based on the one or more sum relevancies.

7. The one or more computer-readable media of claim 6, wherein the one or more sum relevancies comprise a sum relevancy that is calculated based on relevancies between each of word segments and product categories corresponding to the selected one or more nodes, the word segments being generated based on the product information.

8. The one or more computer-readable media of claim 6, wherein the one or more sum relevancies comprise a sum relevancy that is calculated based on relevancies between each of one or more phrases and the product categories corresponding to the selected one or more nodes, the one or more phrases being generated based on the word segments.

9. The one or more computer-readable media of claim 1, wherein the acts further comprise:

setting the node as a parent node; and

determining one or more child nodes of the parent node.

10. A computer-implemented method comprising:

receiving product information;

forming a phrase including one or more word segments that are generated based on the product information;

determining one or more nodes of a product category tree, the one or more nodes connecting with one parent node and having relevancies associated with the phrase greater than a predetermined threshold; and

determining a node of the one or more nodes based on relevancies between each of the one or more word segments and product categories corresponding to the one or more nodes.

11. The computer-implemented method of claim 10, wherein the determining the node of the one or more nodes comprises determining the node of the one or more nodes based further on each of one or more phrases and the product categories corresponding to the one or more nodes, the one or more phrases being generated based on the one or more word segments.

12. The computer-implemented method of claim 10, further comprising:

determining that the node has one or more leaf nodes;

determining a leaf node of the one or more leaf nodes based on relevancies between the phrase and product categories corresponding to the one or more leaf nodes; and

associating the product information with a product category corresponding the leaf node.

13. The computer-implemented method of claim 10, further comprising:

determining that the node has no leaf node; and

determining one or more child nodes of the node.

14. The computer-implemented method of claim 10, wherein the one parent node is a root node of the product category tree.

15. A computer-implemented method comprising:

determining a node and a parent node of the node that are associated with a product category tree, the node corresponding to a product category that includes multiple product titles;

conducting word segmentation on a product title of the multiple product titles to generate a phrase including one or more word segments; and

calculating a relevancy between the phrase and the product category based on relationship between the phrase and the multiple product titles.

16. The computer-implemented method of claim 15, wherein calculating the relevancy comprises calculating based on a number that the phrase is contained in the multiple product titles.

17. The computer-implemented method of claim 15, wherein calculating the relevancy comprises calculating based on:

a number of the multiple product titles included; and

a number of product titles that are included in a product category corresponding to a parent node of the node.

18. The computer-implemented method of claim 15, wherein calculating the relevancy comprises calculating based on a first weighting value that is calculated based on a number that the phrase is contained in the product category and a number of the multiple product titles included in the product category, and based on a second weighting value that is calculated based on a number that the phrase is contained in the product category and a number of the multiple product titles included in the product category.

19. The computer-implemented method of claim 18, wherein the second weighting value is calculated based further on:

a number that the phrase is contained in a product category corresponding to a parent node of the node; and

a number of product titles that are included in a product category corresponding to a parent node of the node.

20. The computer-implemented method of claim 15, further comprising storing the relevancy in a database associated with a cloud computing platform.

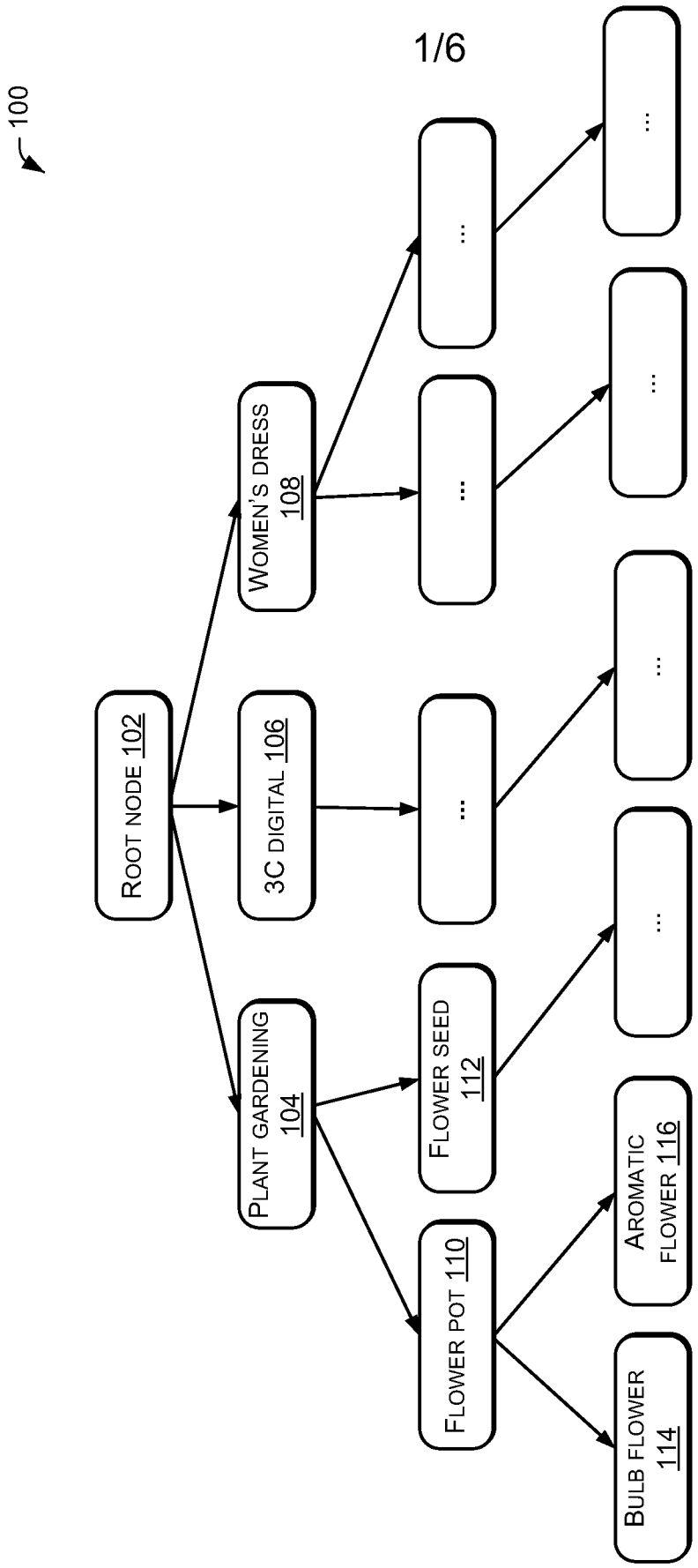

100
ROOT NODE 102
PLANT GARDENING 104
3C DIGITAL 106
WOMEN'S DRESS 108
FLOWER POT 110
FLOWER SEED 112
BULB FLOWER 114
AROMATIC FLOWER 116
...

FIG. 1

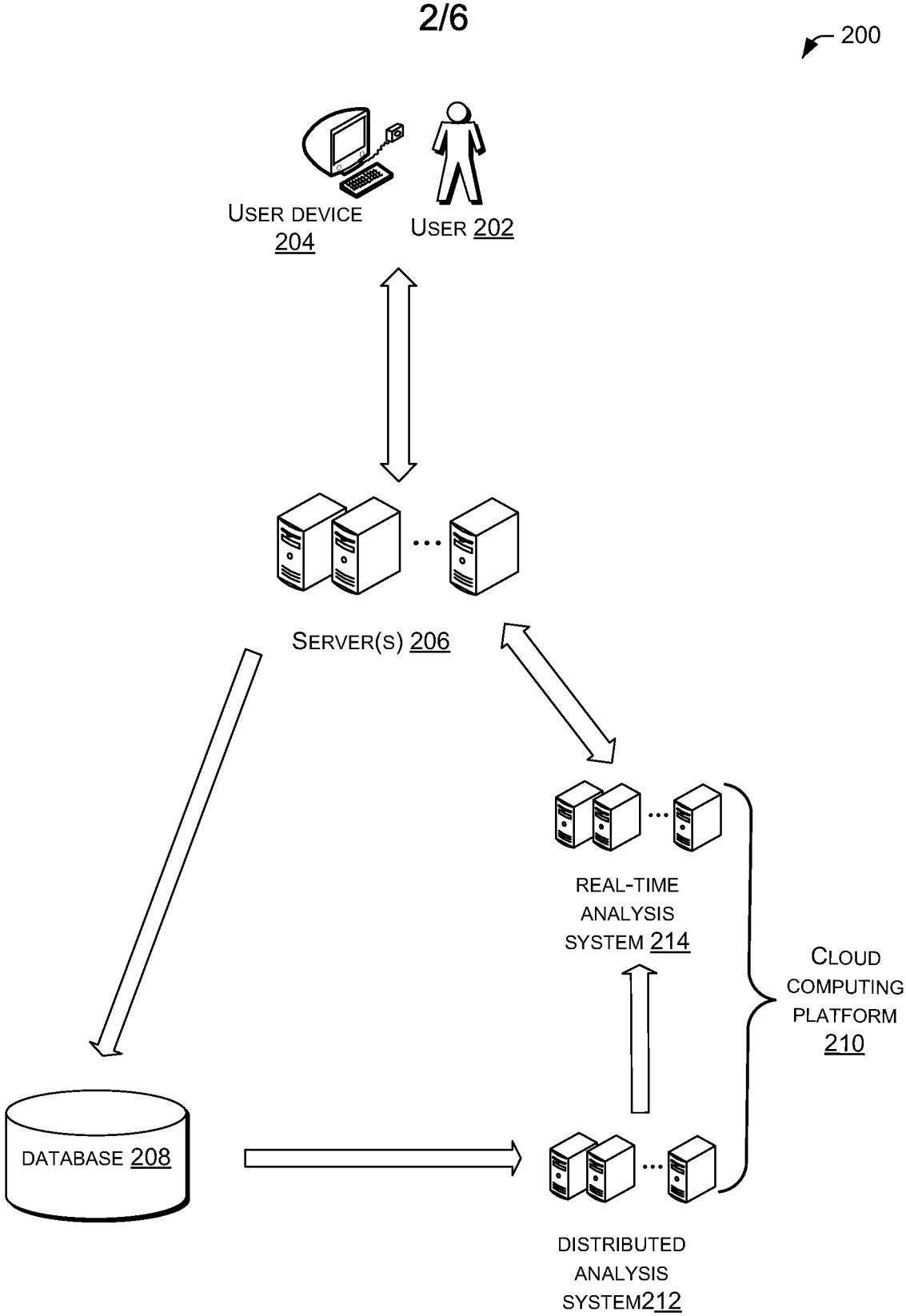

200
USER DEVICE 204
USER 202
SERVER(S) 206
REAL-TIME ANALYSIS SYSTEM 214
CLOUD COMPUTING PLATFORM 210
DATABASE 208
DISTRIBUTED ANALYSIS SYSTEM212

FIG. 2

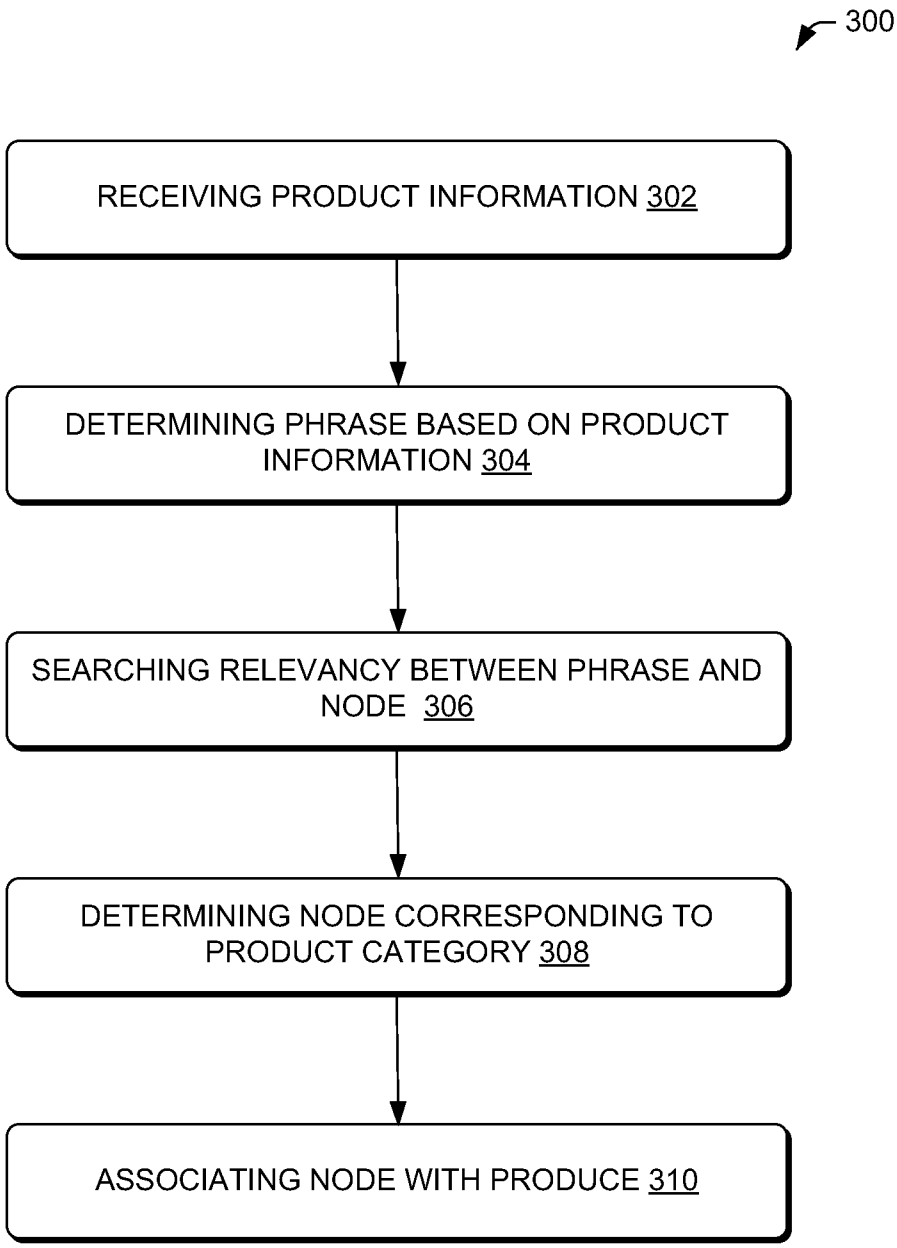

300
RECEIVING PRODUCT INFORMATION 302
DETERMINING PHRASE BASED ON PRODUCT INFORMATION 304
SEARCHING RELEVANCY BETWEEN PHRASE AND NODE 306
DETERMINING NODE CORRESPONDING TO PRODUCT CATEGORY 308
ASSOCIATING NODE WITH PRODUCE 310

FIG. 3

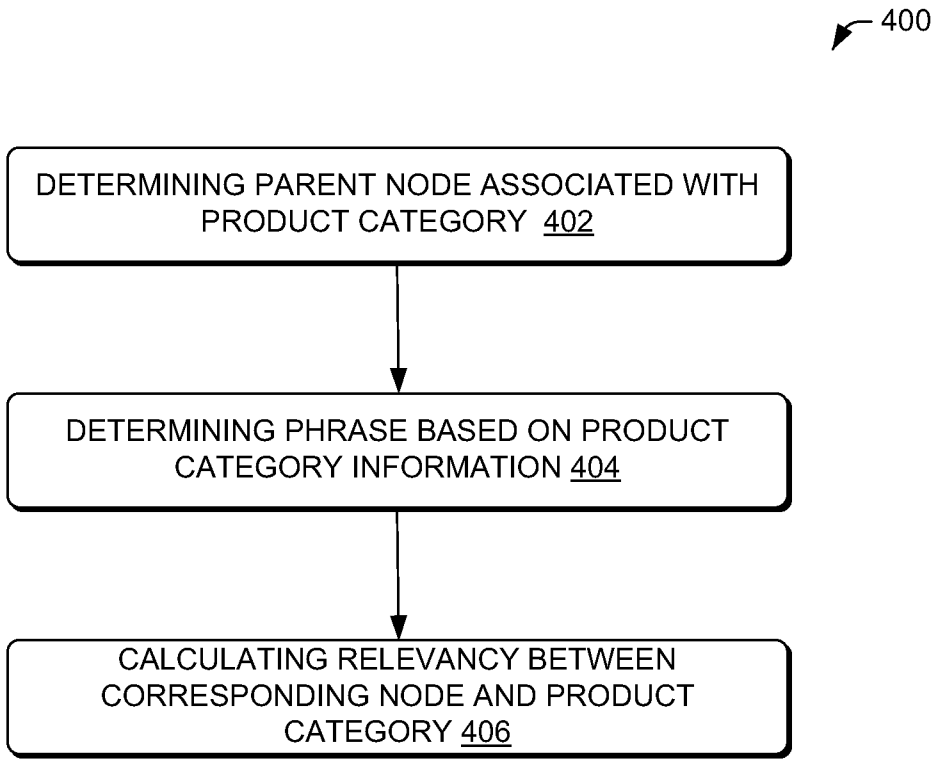

400
DETERMINING PARENT NODE ASSOCIATED WITH PRODUCT CATEGORY 402
DETERMINING PHRASE BASED ON PRODUCT CATEGORY INFORMATION 404
CALCULATING RELEVANCY BETWEEN CORRESPONDING NODE AND PRODUCT CATEGORY 406

FIG. 4

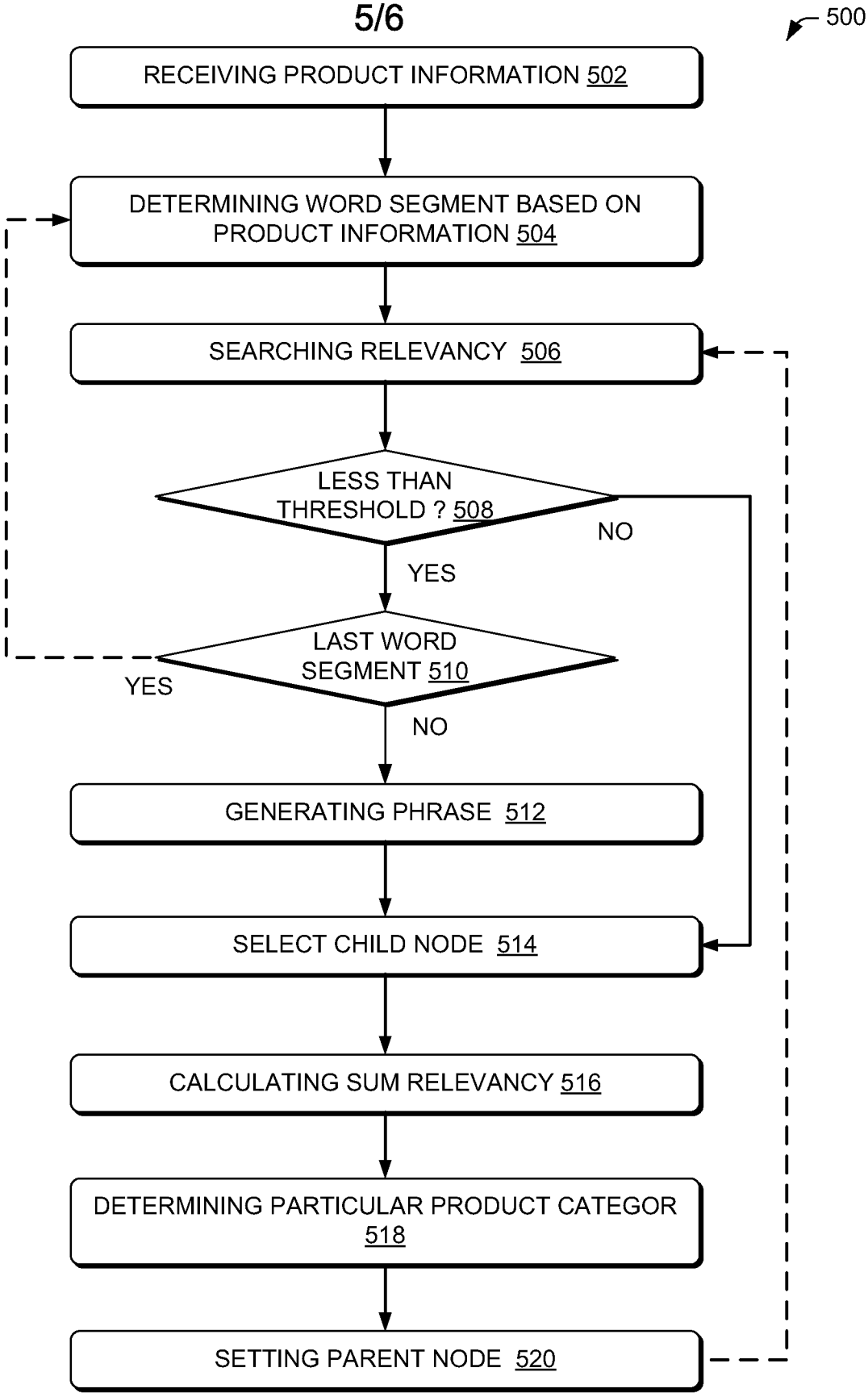

500
RECEIVING PRODUCT INFORMATION 502
DETERMINING WORD SEGMENT BASED ON PRODUCT INFORMATION 504
SEARCHING RELEVANCY 506
LESS THAN THRESHOLD ? 508
NO
YES
LAST WORD SEGMENT 510
YES
NO
GENERATING PHRASE 512
SELECT CHILD NODE 514
CALCULATING SUM RELEVANCY 516
DETERMINING PARTICULAR PRODUCT CATEGOR 518
SETTING PARENT NODE 520

FIG. 5

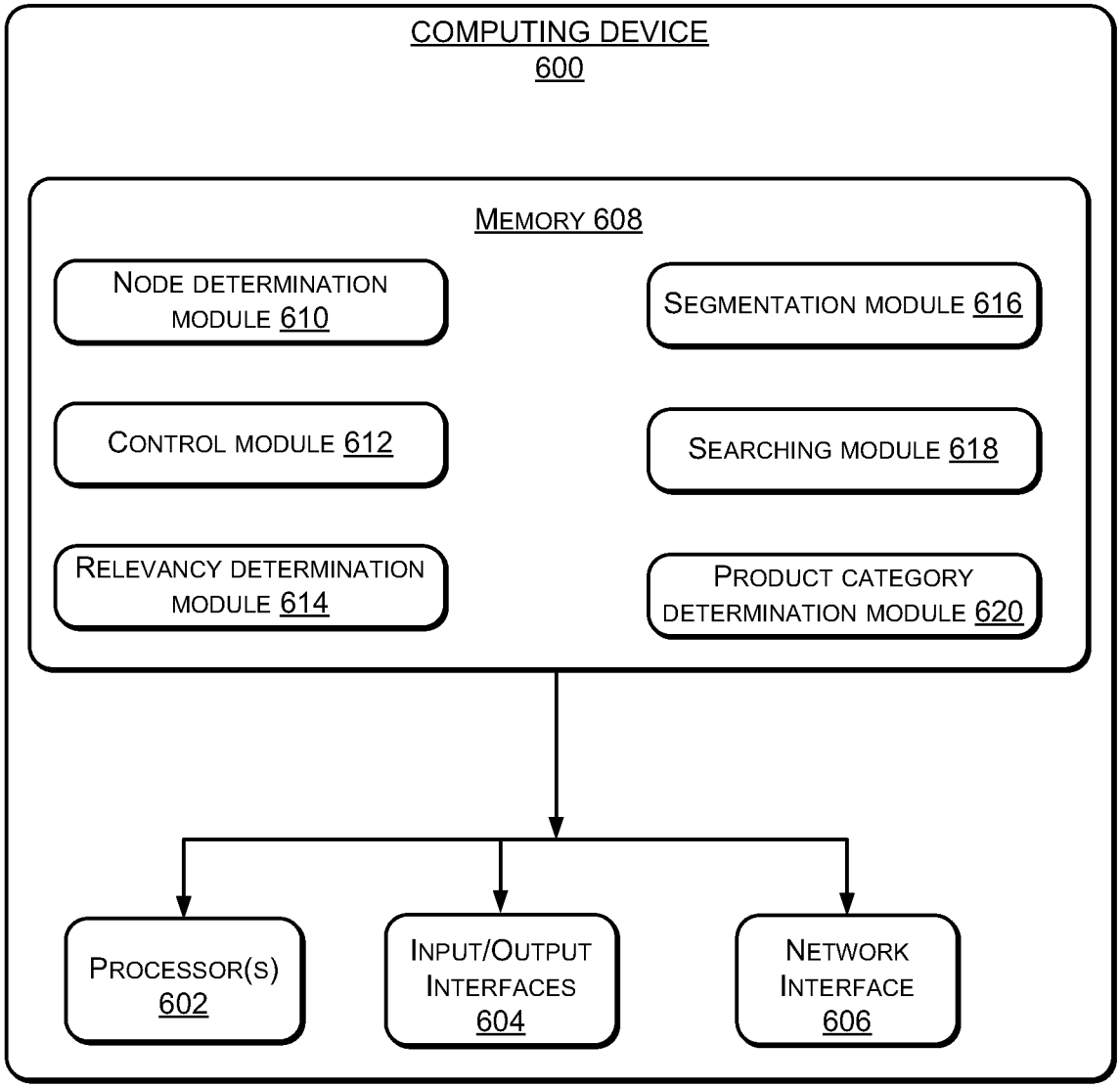

COMPUTING DEVICE
600
MEMORY 608
NODE DETERMINATION MODULE 610
SEGMENTATION MODULE 616
CONTROL MODULE 612
SEARCHING MODULE 618
RELEVANCY DETERMINATION MODULE 614
PRODUCT CATEGORY DETERMINATION MODULE 620
PROCESSOR(S) 602
INPUT/OUTPUT INTERFACES 604
NETWORK INTERFACE 606

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/31875

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 7/00 (2012.01)
USPC - 707/740
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)
IPC(8): G06F 7/00 (2012.01)
USPC: 707/740

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
IPC(8): G06F 7/00 (2012.01) (search term limited - see terms)
USPC: 707/705,736-741; 706/12-14 (search term limited - see terms)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PUBWEST (PGPB, USPT, EPAB, JPAB); Google Patents and Scholar. Search Terms: categorize, product, goods, service, automatic, automated, computer-controlled, tree, root, node, hierarchy, extract, parse, analyze, feature, word, phrase, word segments, calculate, compute, determine, score, relevancy relationship, metric, taxonomy, document, frequency

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 7,870,039 B1 (DOM et al) 11 January 2011 (11.01.2011), entire document; especially Abstract; col 4, ln 16-19; col 6, ln 32-37; col 6, ln 59 - col 7, ln 20; col 10, ln 9-36; col 11, ln 29-41; col 12, ln 21-58; col 16, ln 34-50; col 16, ln 55- col 17, ln 14; col 20, ln 27-50; and Fig. 6	1 - 20
A	US 7,617,200 B2 (Budzik et al.) 10 November 2009 (10.11.2009) entire document	1-20

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

* 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	Date of mailing of the international search report
30 June 2012 (30.06.2012)	06 JUL 2012
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)