



- (51) **International Patent Classification:**
G06F 17/00 (2006.01) G06F 17/30 (2006.01)
- (21) **International Application Number:**
PCT/US201 1/025030
- (22) **International Filing Date:**
16 February 2011 (16.02.2011)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant (for all designated States except US):** HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P. [US/US]; 11445 Compaq Center Drive West, Houston, Texas 77070 (US).
- (72) **Inventors; and**
- (75) **Inventors/ Applicants (for US only):** LUKOSE, Rajan [US/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). SAYERS, Craig P. [US/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). SCHOLZ, Martin B. [DE/US]; 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US).

- (74) **Agents:** CHANG, Marcia et al; Hewlett-Packard Company, Intellectual Property Administration, 3404 East Harmony Road, Mail Stop 35, Fort Collins, Colorado 80528 (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, RO, RS, RU, 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, 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).

[Continued on nextpage]

(54) **Title:** POPULATION CATEGORY HIERARCHIES

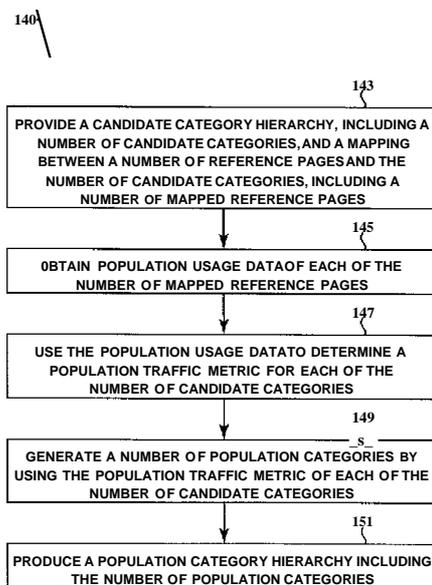


Fig. 1

(57) **Abstract:** Developing a population category hierarchy can include providing a candidate category hierarchy, including a number of candidate categories, and a mapping between a number of reference pages and the number of candidate categories, including a number of mapped reference pages (143). Population usage data of the number of mapped reference pages can be obtained and used to determine a population traffic metric for each of the number of candidate categories (147). A number of population categories can be generated by using the population traffic metric of each of the number of candidate categories (149); and, a population category hierarchy can be produced including the number of population categories (151).

WO 2012/112149 A1

Declarations under Rule 4.17:

- *as to the identity of the inventor (Rule 4.17(i))*
- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(H))*

Published:

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

Population Category Hierarchies

Background

[0001] A user's web browsing history is a rich data source representing a user's implicit and explicit interests and intentions. Completed, recurring, and ongoing tasks of varying complexity and abstraction can be found in a user's web browsing history and is consequently a valuable resource. Mechanisms that organize a user's web browsing history have been introduced. As the internet continues to become ever more essential and the key tool for information seeking and retrieval, various web browsing mechanisms that organize a user's web browsing history have been introduced.

Brief Description of the Drawings

[0002] Figure 1 illustrates an example of a method for developing a population category hierarchy according to the present disclosure.

[0003] Figure 2 illustrates a block diagram of an example of a system for developing a population category hierarchy according to the present disclosure.

[0004] Figure 3 illustrates a block diagram of an example of a computer-readable medium (CRM) in communication with processing resources for developing a population category hierarchy according to the present disclosure.

[0005] Figure 4 illustrates a block diagram of an example of a candidate category hierarchy and a mapping according to the present disclosure.

[0006] Figure 5 illustrates a block diagram of an example of a population category hierarchy and a population mapping according to the present disclosure.

[0007] Figure 6 illustrates an example of a population mapping according to the present disclosure.

Detailed Description

[0008] The present disclosure provides methods, machine readable media, and systems for developing population category hierarchies. A mapping can be provided between a number of reference pages and the number of

candidate categories to create a number of mapped reference pages. Population usage data of the number of mapped reference pages can be obtained and used to determine a population traffic metric for each of the number of candidate categories. A number of population categories can be generated by using the population traffic metric of each of the number of candidate categories; and, a population category hierarchy can be produced, including the number of population categories.

[0009] For example, implementation of a category hierarchy started in the United States can include a category labeled "sports" which can include a subcategory "baseball" but not a subcategory "Australian rules football." If the category path system were ported to Australia it could be advantages for information to be generated that indicates "Australian rules football" should be a subcategory. In an example, the use of web based services such as, for example, Wikipedia™ can be used to tailor a category hierarchy to a population by the use of usage data from a defined population to properly weight categories and subcategories. Additional categories can consume resources and add complexity, but additional tailored categories can also aid in organizing and categorizing hierarchies.

[0010] In the present disclosure, reference is made to the accompanying drawings that form a part hereof, and in which is shown by way of illustration how one or more examples of the disclosure can be practiced. These examples are described in sufficient detail to enable those of ordinary skill in the art to practice the examples of this disclosure, and it is to be understood that other examples can be used and that process, electrical, and/or structural changes can be made without departing from the scope of the present disclosure.

[0011] The figures herein follow a numbering convention in which the first digit corresponds to the drawing figure number and the remaining digits identify an element or component in the drawing. Similar elements or components between different figures can be identified by the use of similar digits. For example, 214 can reference element "14" in Figure 2, and a similar element can be referenced as 314 in Figure 3. Elements shown in the various figures herein can be added, exchanged, and/or eliminated so as to provide a number of

additional examples of the present disclosure. In addition, the proportion and the relative scale of the elements provided in the figures are intended to illustrate the examples of the present disclosure, and should not be taken in a limiting sense.

[0012] Figure 1 is a flow chart illustrating an example of a method 140 for developing a population category hierarchy according to one or more examples of the present disclosure. Examples of the present disclosure are not limited to the steps illustrated in Figure 1. The method includes a candidate category hierarchy, including a number of candidate categories, and a mapping provided at step 143. A candidate category hierarchy can include a publicly available source of ontology information in which various concepts are assigned to one or more categories. A candidate category hierarchy can be provided, for example, by a web-based service. For example, in the Wikipedia™ database, each of the articles is assigned a particular concept. In addition, the concepts are assigned to particular categories and sub-categories defined by the editors of the Wikipedia™ database. Such examples can be beneficial in providing for a starting point for development of a population category hierarchy.

[0013] According to an example, some or all of the provided candidate category hierarchy can be manually defined. The candidate category levels that are not manually defined can be computed from categorical information contained in a labeled text data source. A labeled text data source generally comprises a third-party database of articles such as Wikipedia™, Freebase™, IMDb™, among others. For instance, a user can define a category and one or more subcategories and can rely on the candidate category levels contained in the labeled text data source for the remaining subcategories in the hierarchy of predefined category levels. According to an example, a user can define the hierarchy of predefined candidate category levels as a tree structure and can map the categories of the labeled text data source into the tree structure. A tree structure is a common way that can be used to represent a hierarchical nature of a structure in a graphical format. An example can include a recorded relevance of each concept to each category as the probability that another article that mentions that concept would appear in that category.

[0014] A mapping is provided between the candidate categories in the candidate category hierarchy and a number of reference pages, including a number of mapped reference pages at step 143. A reference page can include, for example, in the Wikipedia™ database, the articles assigned to concepts which are additionally assigned to particular categories and subcategories. A reference page could further, for example, be manually defined. Construction of a mapping between the number of reference pages and the candidate categories can be accomplished by a computing device, such as the one discussed below and shown in figure 2. For example, the labeled text data source corpus can be analyzed by the computing device to find categories for each concept by mapping the labeled text data source categories into a category graph (e.g., a manually constructed category tree), find phrases related to each category by using the text of reference pages assigned to concepts of each category, find phrases related to each concept by using text anchor tags which point to that concept, and evaluate counts of occurrences to determine the probability that an occurrence of a particular phrase indicated by the text is relevant to a particular category or a particular concept. For example, if 10% of reference pages (e.g., articles) containing the text "Tiger" are in the category "Golf, then the probability of an input text being in the category "Golf, given that it contains the text "Tiger", is 0.1. As another example, if 30% of the occurrences of the text "Tiger" link to the article labeled with the concept "Tiger Woods", then the probability that the input text is related to "Tiger Woods", given that we've observed it contains the text "Tiger", is 0.3.

[0015] Population usage data of each of the number of mapped reference pages are obtained at step 145. A population can be defined, for example, by a particular geographic region or politically defined country borders. A population can further be defined, for example, by type of device used to access the reference page (e.g., mobile device). In an example, a population can be defined by a number of resources including: user defined parameters, cell phone signals, cell phone tower signals, Global Position System (GPS) device signals, Internet Service Provider (ISP) information, and web-based service information. Web-based services can include Wikipedia™, Freebase™,

IMDB™, among others. Population usage data can include timestamps, which can be used to estimate a time each visitor spent on each mapped reference page, and a total number of visits to each mapped reference page. Timestamps can, for example, be visitor specific and include a time the visitor arrives at a page and a time the visitor leaves the page. In such examples, the number of timestamps can be considered a number of sets of timestamps (e.g., beginning and ending timestamps). The total number of visits can, for example, be a total number of visits to a page per visitor (e.g., the number of times a visitor frequents a page). In another example, the total number of visits can be the overall total of the number of visits to a reference page. In another example, population usage data can include usage data of pages related to the number of mapped reference pages. Such examples can be beneficial in providing greater information regarding a defined population and tailoring categories to aid in organization and categorization of hierarchies.

[0016] A population traffic metric is determined for each of the number of candidate categories by using the population usage data at step 147. A population traffic metric can include, for example, a summation of the number of visits to each of the number of reference pages mapped to the candidate category. In an example, a population traffic method can be determined by assigning a weight according to the timestamp population usage data. For example, the weight given to the population traffic metric can have a positive relationship to the estimated time each visitor spent on each mapped reference page.

[0017] The population traffic metric of each of the number of candidate categories is used to generate a number of population categories at step 149. In an example, candidate categories for which the population traffic metric falls below a chosen threshold level can be deleted to create the number of population categories. Population categories can, for example, be generated through rank of candidate categories by population traffic metric and retaining candidate categories above a threshold. In another example, addition of candidate categories can generate the number of population categories. Population categories can be generated through a merge and/or split of

candidate categories. A merge and/or split can, for example, be done according to an input from a user. For example, candidate categories with population traffic metrics lower than a threshold to justify designation of the candidate category as an entire population category, but higher than a threshold to justify deletion of the candidate category could be merged to obtain a threshold population traffic metric to justify a population category. In another example, a candidate category can have a population traffic metric above a threshold at which the population traffic metric allows adequate categorization. Consequently, the candidate category can, for example, be split into multiple, population categories.

[0018] At step 151, a population category hierarchy is produced that includes the number of population categories. In an example, the mapping can be modified to create a population mapping, wherein the number of mapped reference pages are re-mapped to the number of population categories. In another example, the population mapping can be created by the use of relevance metrics for each of the number of population categories to identify a relevance level for each of the number of population categories and the number of mapped reference pages. The mapped reference pages can, for example, be re-mapped according to the relevance level of each of the number of population categories. For example, a relevance metric can take into account the number of times a mapped reference page is mapped in the population mapping. A user, in an example, can provide input to approve or disapprove of a produced population mapping. Such user input examples can be beneficial because a user can take into account certain factors that the produced population mapping did not in a particular instance (e.g., time of year or current trends) and allow the user to decide to disapprove and/or alter the mapping. Another example includes repetition of any of the above steps at a defined interval to update the number of population categories.

[0019] Figure 2 illustrates a block diagram 200 of an example of a computer-readable medium (CRM) 220 in communication with a computing device 212 (e.g., Java application server) having memory resources 217 and processor resources of more or fewer than 214-1, 214-2, 214-3, that can be in

communication with, and/or receive a tangible non-transitory computer readable medium (CRM) 220 storing a set of computer readable instructions 215 executable by one or more of the processor resources (e.g., 214-1, 214-2, 214-3) for profiling a server, as described herein.

[0020] Memory resources 217 can include volatile and/or non-volatile memory. Volatile memory, as used herein, can include memory that depends upon power to store information, such as various types of dynamic random access memory (DRAM), among others. Non-volatile memory, as used herein, can include memory that does not depend upon power to store information. Examples of non-volatile memory can include solid state media such as flash memory, EEPROM, phase change random access memory (PCRAM), magnetic memory such as a hard disk, tape drives, floppy disk, and/or tape memory, optical discs, digital video discs (DVD), high definition digital versatile discs (HD DVD), compact discs (CD), and/or a solid state drive (SSD), flash memory, etc., as well as other types of machine-readable media.

[0021] Processor resources can execute computer-readable instructions 215 that are stored on an internal or external non-transitory computer-readable medium 220. A non-transitory computer-readable medium (e.g., computer readable medium 220), as used herein, can include volatile and/or non-volatile memory.

[0022] The non-transitory computer-readable 220 medium can be integral, or communicatively coupled, to a computing device, in either in a wired or wireless manner. For example, the non-transitory computer-readable medium can be an internal memory, a portable memory, a portable disk, or a memory located internal to another computing resource (e.g., enabling the computer-readable instructions to be downloaded over the Internet).

[0023] The CRM 220 can be in communication with the processor resources (e.g., 214-1, 214-2, 214-3) via a communication path 276. The communication path 276 can be local or remote to a machine associated with the processor resources (214-1, 214-2, 214-3). Examples of a local communication path 276 can include an electronic bus internal to a machine such as a computer where the CRM 220 is one of volatile, non-volatile, fixed,

and/or removable storage medium in communication with the processor resources (e.g., 214-1, 214-2, 214-3) via the electronic bus. Examples of such electronic buses can include Industry Standard Architecture (ISA), Peripheral Component Interconnect (PCI), Advanced Technology Attachment (ATA), Small Computer System Interface (SCSI), Universal Serial Bus (USB), among other types of electronic buses and variants thereof.

[0024] In other examples, the communication path 276 can be such that the CRM 220 is remote from the processor resources (e.g., 214-1, 214-2, 214-3) such as in the example of a network connection between the CRM 220 and the processor resources (e.g., 214-1, 214-2, 214-3). That is, the communication path 276 can be a network connection. Examples of such a network connection can include a local area network (L-AN), a wide area network (WAN), a personal area network (PAN), and the Internet, among others. In such examples, the CRM 220 may be associated with a first computing device and the processor resources (e.g., 214-1, 214-2, 214-3) may be associated with a second computing device 212 (e.g., a Java application server).

[0025] Figure 3 illustrates a block diagram of an example of a computing system 300 for developing a population category hierarchy according to the present disclosure. However, examples of the present disclosure are not limited to a particular computing system configuration. The system 300 can include processor resources 314 and memory resources (e.g., volatile memory 316 and/or non-volatile memory 318) for executing instructions stored in a tangible non-transitory medium (e.g., volatile memory 316, non-volatile memory 318, and/or computer-readable medium 320) and/or an application specific integrated circuit (ASIC) including logic configured to perform various examples of the present disclosure. A computer (e.g., a computing device) can include and/or receive a tangible non-transitory computer-readable medium 320 storing a set of computer-readable instructions (e.g., software) via an input device 322. In an example, the input device 322 can receive input from a number of web based services 373. As used herein, processor resources 314 can include one or a plurality of processors such as in a parallel processing system. Memory resources can include memory addressable by the processor resources 314 for

execution of computer-readable instructions. The computer-readable medium 320 can include volatile and/or non-volatile memory such as random access memory (RAM), magnetic memory such as a hard disk, floppy disk, and/or tape memory, a solid state drive (SSD), flash memory, phase change memory, etc. In some examples, the non-volatile memory 318 can be a database including a plurality of physical non-volatile memory devices. In various examples, the database can be local to a particular system or remote (e.g., including a plurality of non-volatile memory devices 318).

[0026] The processor resources 314 can control the overall operation of the system 300. The processor resources 314 can be connected to a memory controller 324, which can read and/or write data from and/or to volatile memory 316 (e.g., RAM). The memory controller 324 can include an ASIC and/or a processor with its own memory resources (e.g., volatile and/or non-volatile memory). The volatile memory 316 can include one or a plurality of memory modules (e.g., chips).

[0027] The processor resources 314 can be connected to a bus 326 to provide for communication between the processor resources 314, and other portions of the system 300. The non-volatile memory 318 can provide persistent data storage for the system 300. The graphics controller 328 can connect to a user interface 330, which can provide an image to a user based on activities performed by the system 300.

[0028] Each system can include a computing device including control circuitry such as a processor, a state machine, application specific integrated circuit (ASIC), controller, and/or similar machine. As used herein, the indefinite articles "a" and/or "an" can indicate one or more than one of the named object. Thus, for example, "a processor" can include one processor or more than one processor, such as a parallel processing arrangement.

[0029] The control circuitry can have a structure that provides a given functionality, and/or execute computer-readable instructions that are stored on a non-transitory computer-readable medium (e.g. non-transitory computer-readable medium 320). The non-transitory computer-readable medium can be integral, or communicatively coupled, to a computing device, in either in a wired

or wireless manner. For example, the non-transitory computer-readable medium 320 can be an internal memory, a portable memory, a portable disk, or a memory located internal to another computing resource (e.g., enabling the computer-readable instructions to be downloaded over the Internet). The non-transitory computer-readable medium 320 can have computer-readable instructions 315 stored thereon that are executed by the control circuitry (e.g., processor) to provide a particular functionality.

[0030] The non-transitory computer-readable medium, as used herein, can include volatile and/or non-volatile memory. Volatile memory can include memory that depends upon power to store information, such as various types of dynamic random access memory (DRAM), among others. Non-volatile memory can include memory that does not depend upon power to store information. Examples of non-volatile memory can include solid state media such as flash memory, EEPROM, phase change random access memory (PCRAM), among others. The non-transitory computer-readable medium can include optical discs, digital video discs (DVD), Blu-Ray Discs, compact discs (CD), laser discs, and magnetic media such as tape drives, floppy discs, and hard drives, solid state media such as flash memory, EEPROM, phase change random access memory (PCRAM), as well as other types of computer-readable media.

[0031] Figure 4 illustrates a block diagram of an example of a candidate category hierarchy 460 and a mapping 464 according to the present disclosure. Candidate category hierarchy 460 and the mapping 464 can, for example, be provided as described above (e.g., step 143 in Figure 1). A candidate category hierarchy 460 can, for example, contain candidate categories 462. In an example, candidate categories 462 can be linked to a number of other candidate categories 462. For example, a candidate category 462 labeled "Sports" can be linked to other candidate categories 462 labeled, for example, "Baseball," "Football," and "Golf." Candidate categories 462, can be mapped to reference pages 466. Once a candidate category 462 has been mapped to a reference page 466, the reference page 466 is referred to as a mapped reference page 468. The block diagram illustrated in Figure 4 includes a number of reference pages 466 that are not mapped to candidate categories

462 and a number of mapped reference pages 468 that are mapped to candidate categories 462. Candidate categories can be mapped to reference pages 466 to create a mapping 464 that contains a number of mapped reference pages 468. For example, a candidate category 462 labeled "Baseball" can be mapped to reference pages 466 such as the Minnesota Twins baseball club homepage (e.g., www.minnesota.twins.mlb.com), the official website of Major League baseball (e.g., www.mlb.com), or the Wikipedia article on baseball (e.g., www.en.wikipedia.org/wiki/Baseball) to form a number of mapped reference pages 468.

[0032] Each mapped reference page 468, in an example, contains population usage data 470, however examples are not so limited as one or more mapped reference pages 468 may not include population usage data 470 and/or a reference page 466 (e.g., an unmapped reference page) may include population usage data 470. Population usage data 470 can include a number of timestamps 478 and/or a total number of visits 480 to a particular mapped reference page 468. In an example, the population usage data 470 can be stipulated by a number of parameters, including, but not limited to: user defined parameters 477, cell phone signals 479, cell phone tower signals 481, GPS device signals 483, ISP information 485, and/or web-based service information 487. For example, the population usage data 470 can be stipulated by a user defined parameter 477 to contain data only regarding visitors from a specific region (e.g., Australia, Europe, North America, etc.). In an example, multiple parameters can be used to stipulate the population usage data 470.

[0033] Figure 5 illustrates a block diagram of an example of a population category hierarchy 576 and a population mapping 592 according to the present disclosure. Population category hierarchy 576 can, for example, be produced as described above (e.g., step 151 in figure 1). Prior to the population mapping 592, the population usage data 570 of each mapped reference page 568 can be used to determine a population traffic metric for each candidate category to generate a number of population categories 574. For example, a user can define a lower threshold limit for a population traffic metric. In an example, the lower threshold can be a specified total number of visits (e.g., visits 480

illustrated in Figure 4) to all mapped reference pages 568 of a candidate category (e.g., candidate category 462 illustrated in Figure 4). In such an example, any candidate category, below the specified total number of visits, can be removed from the mapping (e.g., mapping 464 illustrated in Figure 4) to create population categories 574 in a population mapping 592.

[0034] Population categories 574 of population mapping 592 can maintain the same mapped reference pages as the candidate category or categories from which they were created. For example, if the candidate category "Minor League Baseball Teams in Minnesota" was mapped to the mapped reference page "St. Paul Saints" and the candidate category "Minor League Baseball Teams in Minnesota" did not receive above the minimum threshold of a total number of visits for all mapped reference pages (e.g., "St. Paul Saints") mapped to the candidate category, the candidate category can be merged with a similar candidate category "Minor League Baseball Teams in the USA" to create a population category 574 "Minor League Baseball Teams in the USA" that maintained the mapping to the mapped reference page 568 "St. Paul Saints" as well as any existing mapped reference pages of the candidate category "Minor League Baseball Teams in the USA." In such an example, the mapped reference pages 568 are maintained (e.g., not re-mapped as discussed in regards to Figure 6) whereas the candidate categories are manipulated to form population categories 574.

[0035] The population categories 574, in an example, can be organized such that a population category 574 can be linked to other population categories 574 to create a population category hierarchy 576. In another example, population usage data 570 of related pages 588 to the number of mapped reference pages 568 can be used to produce the population category hierarchy 576. For example, a mapped reference page 568 can be the Minnesota Twins baseball club homepage, "www.minnesota.twins.mlb.com." Related pages 588 can include players on the Minnesota Twins baseball club (e.g., www.joe-mauer.org, www.thisisdspan.com, or www.twitter.com/mcuddy5) or blogs about the Minnesota Twins baseball club (e.g., www.aarongleeman.com, www.nickstwinsblog.com, or www.twinkietown.com). Such related pages 588

examples can be beneficial because the use of related pages 588 can make the population traffic metric more tailored to a particular topic and consequently can lead to the creation of a population category hierarchy 576 and population mapping 592 that are more reflective of a particular population.

[0036] Figure 6 illustrates an example of a population mapping 692 according to the present disclosure. A population mapping can be created by re-mapping the number of mapped reference pages 668 to form the number of population categories 674 (e.g., after step 151 in Figure 1). For example, a mapped reference page 668 referencing Tiger Woods can be mapped to a candidate category "Golf" (e.g., candidate category 462 illustrated in Figure 4). The mapped reference page referencing Tiger Woods 668 can, in an example, be re-mapped, based on the population usage data 670, for example, to candidate categories "Celebrities," "Athletes," and/or "Sports Apparel" to create a number of population categories 674 in population mapping 692. Such re-mapping examples can be beneficial to update a population mapping 692 or make a more comprehensive population mapping 692.

[0037] Although specific examples have been illustrated and described herein, those of ordinary skill in the art will appreciate that an arrangement calculated to achieve the same results can be substituted for the specific examples shown. This disclosure is intended to cover adaptations or variations of one or more examples of the present disclosure. It is to be understood that the above description has been made in an illustrative fashion, and not a restrictive one. Combination of the above examples, and other examples not specifically described herein will be apparent to those of skill in the art upon reviewing the above description. The scope of the one or more examples of the present disclosure includes other applications in which the above structures and methods are used. Therefore, the scope of one or more examples of the present disclosure should be determined with reference to the appended claims, along with the full range of equivalents to which such claims are entitled.

[0038] Throughout the specification and claims, the meanings identified below do not necessarily limit the terms, but merely provide illustrative examples for the terms. The meaning of "a," "an," and "the" includes plural reference, and

the meaning of "in" includes "in" and "on." The term "a number of" is meant to be understood as including at least one but not limited to one. The phrase "in an example," as used herein does not necessarily refer to the same example, although it can.

What is claimed:

1. A method for developing a population category hierarchy, comprising:
 - providing a candidate category hierarchy, including a number of candidate categories, and a mapping between a number of reference pages and the number of candidate categories, including a number of mapped reference pages;
 - obtaining population usage data of each of the number of mapped reference pages;
 - using the population usage data to determine a population traffic metric for each of the number of candidate categories;
 - generating a number of population categories by using the population traffic metric of each of the number of candidate categories; and
 - producing a population category hierarchy including the number of population categories.
2. The method of claim 1, wherein the population usage data includes a number of timestamps and a total number of visits to each of the number of mapped reference pages wherein the number of timestamps are used to estimate a time for each of a number of visitors spent on each mapped reference page.
3. The method of claim 2, wherein determining the population traffic metric includes weighting according to the number of timestamps, wherein the weight given and the estimated time each visitor spent on each mapped reference page is a positive relationship.
4. The method of claim 2, wherein determining the population traffic metric includes summing the total number of visits to each of the number of mapped reference pages to a particular one of the number of candidate categories.

5. The method of claim 1, wherein the method further includes modifying the mapping to create a population mapping, wherein the number of mapped reference pages are re-mapped to form the number of population categories to create the population mapping.

6. The method of claim 5, wherein modifying includes:
using relevance metrics for each of the number of population categories to identify a relevance level for each of the number of population categories and the number of mapped reference pages; and
re-mapping the number of mapped reference pages according to the relevance level of each of the number of population categories.

7. The method of claim 1, wherein generating the number of population categories includes reducing or merging candidate categories.

8. The method of claim 1, wherein generating the number of population categories includes adding or splitting candidate categories.

9. The method of claim 1, wherein generating the number of population categories includes:
receiving an input from a user;
merging candidate categories; and
splitting candidate categories.

10. The method of claim 1, wherein the method includes obtaining the candidate category hierarchy and the population usage data from a number of web-based services.

11. The method of claim 1, wherein the method includes defining the population according to at least one of a number of user defined parameters, cell phone signals, cell phone tower signals, Global Positioning Satellite (GPS)

device signals, Internet Service Provider (ISP) information, and web-based service information.

12. A non-transitory computer-readable medium storing a set of instructions executable by a processor to:

- provide a candidate category hierarchy, including a number of candidate categories, and a mapping between a number of reference pages and the number of candidate categories, including a number of mapped reference pages;

- obtain population usage data of each of the number of mapped reference pages;

- use the population usage data to determine a population traffic metric for each of the number of candidate categories;

- generate a number of population categories by using the population traffic metric of each of the number of candidate categories;

- produce a population category hierarchy including the number of population categories; and

- modify the mapping to create a population mapping, wherein the number of mapped reference pages are re-mapped to the number of population categories to create the population mapping.

13. The non-transitory computer-readable medium of claim 12, wherein the medium further includes instructions to use population usage data of pages related to the number of mapped reference pages to produce the population category hierarchy.

14. The non-transitory computer-readable medium of claim 12, wherein the instructions to modify the mapping further include instructions to:

- use relevance metrics for each of the number of population categories to identify a relevance level for each of the number of population categories and the number of mapped reference pages; and

re-map the number of mapped reference pages according to the relevance level of each of the number of population categories.

15. A system for developing a population category hierarchy, comprising:
a computing device including:
a memory; and
a number of processors coupled to the memory, to:
provide a candidate category hierarchy, including a number of candidate categories, and a mapping between a number of reference pages and the number of candidate categories, including a number of mapped reference pages;
obtain population usage data of each of the number of mapped reference pages;
use the population usage data to determine a population traffic metric for each of the number of candidate categories;
generate a number of population categories by using the population traffic metric of each of the number of candidate categories;
produce a population category hierarchy including the number of population categories; and
modify the mapping to create a population mapping, wherein the number of mapped reference pages are re-mapped to the number of population categories to create the population mapping.

1/6

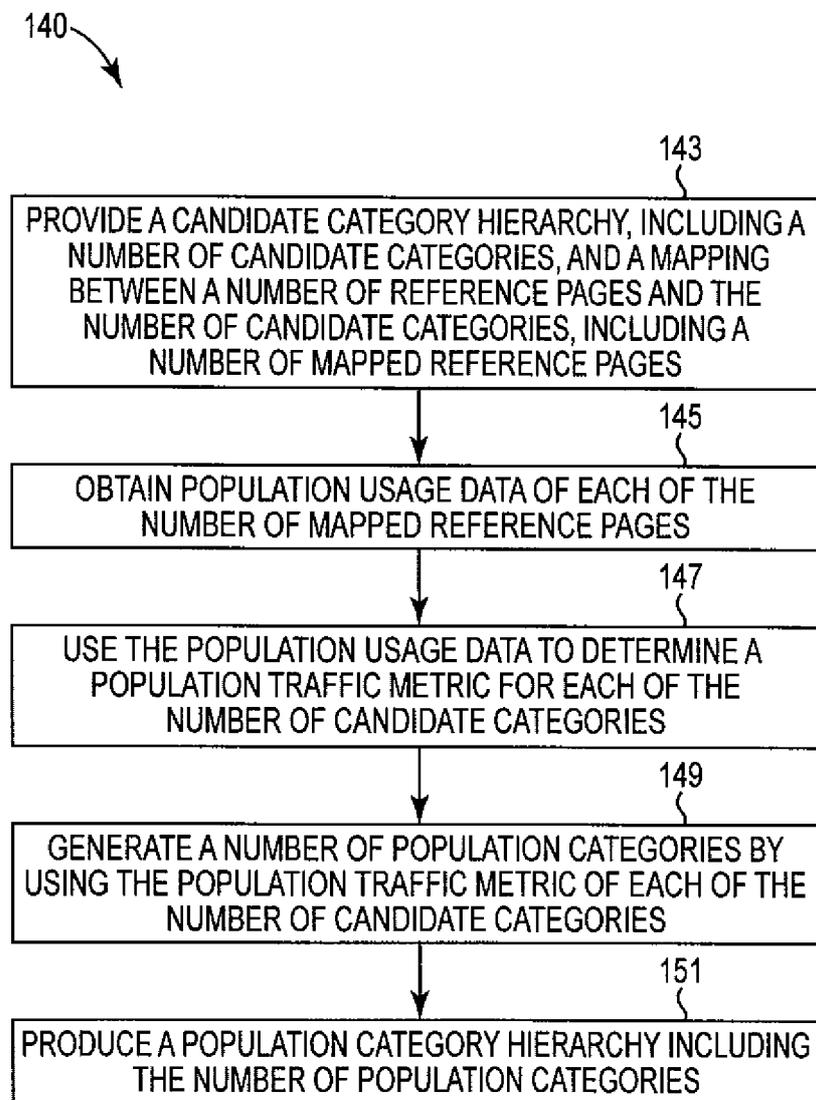


Fig. 1

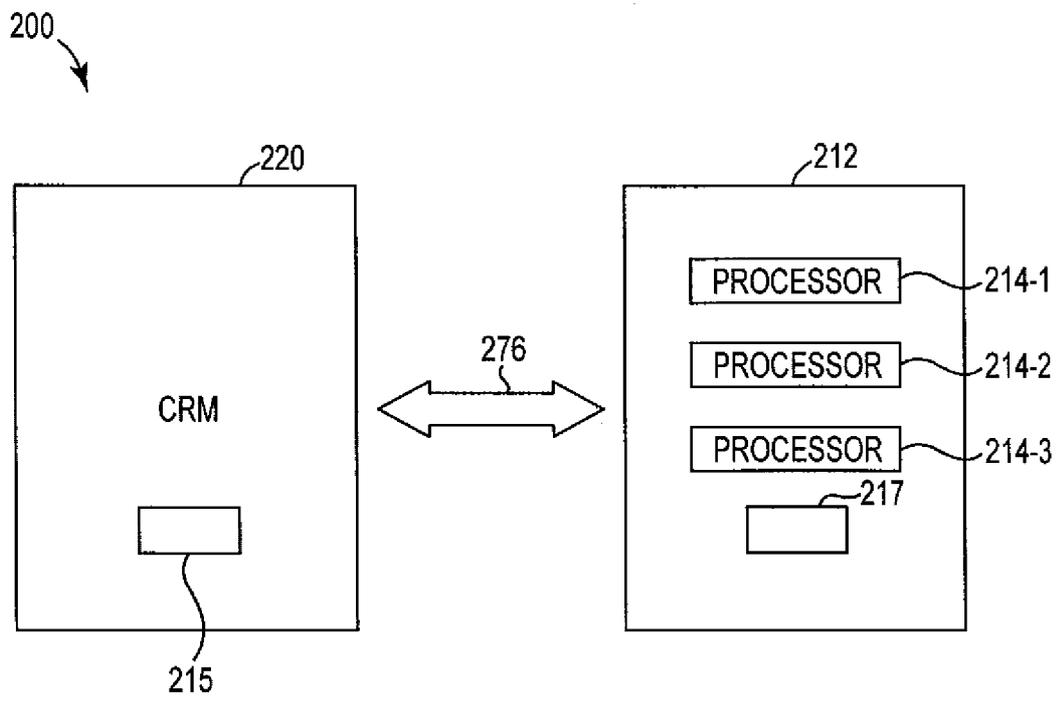


Fig. 2

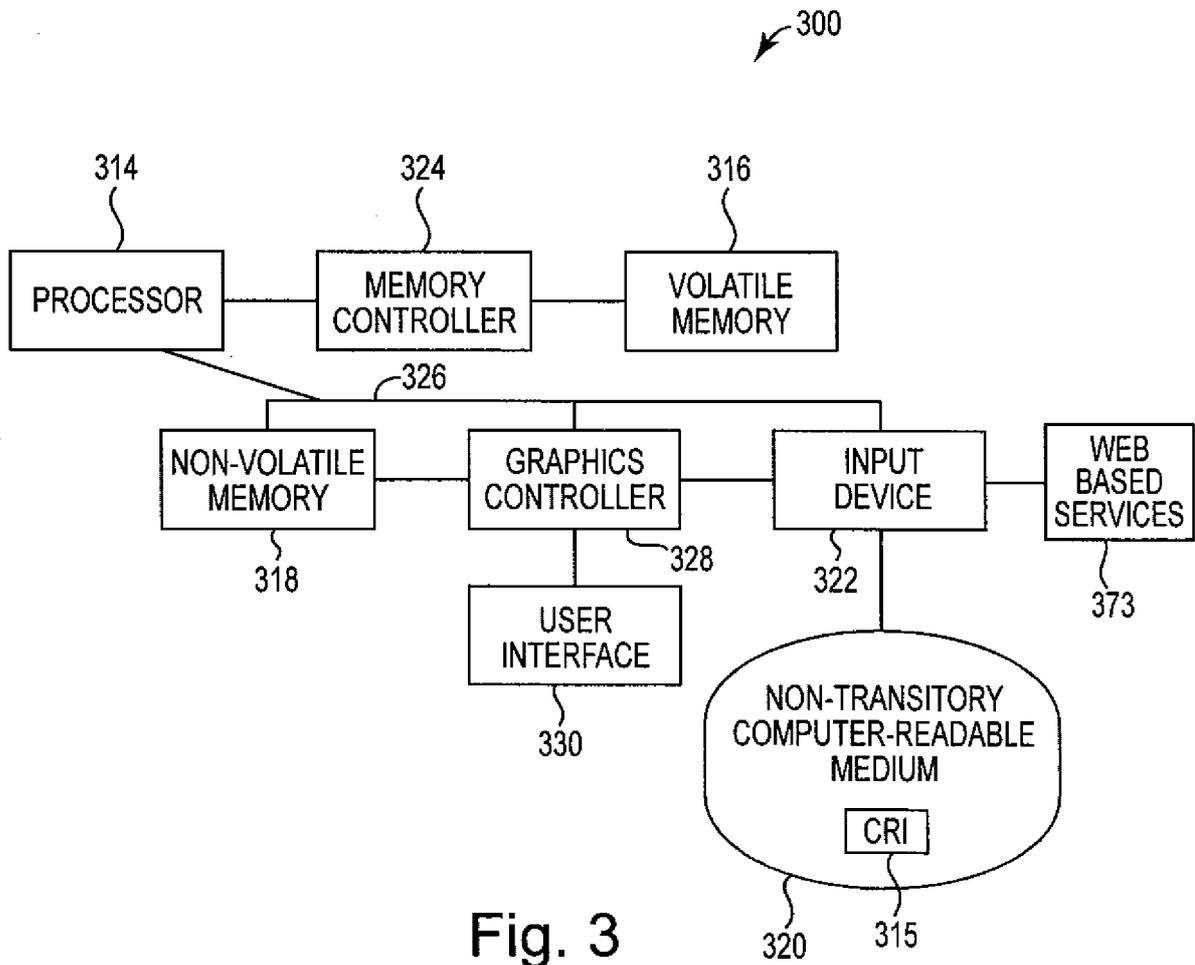


Fig. 3

4/6

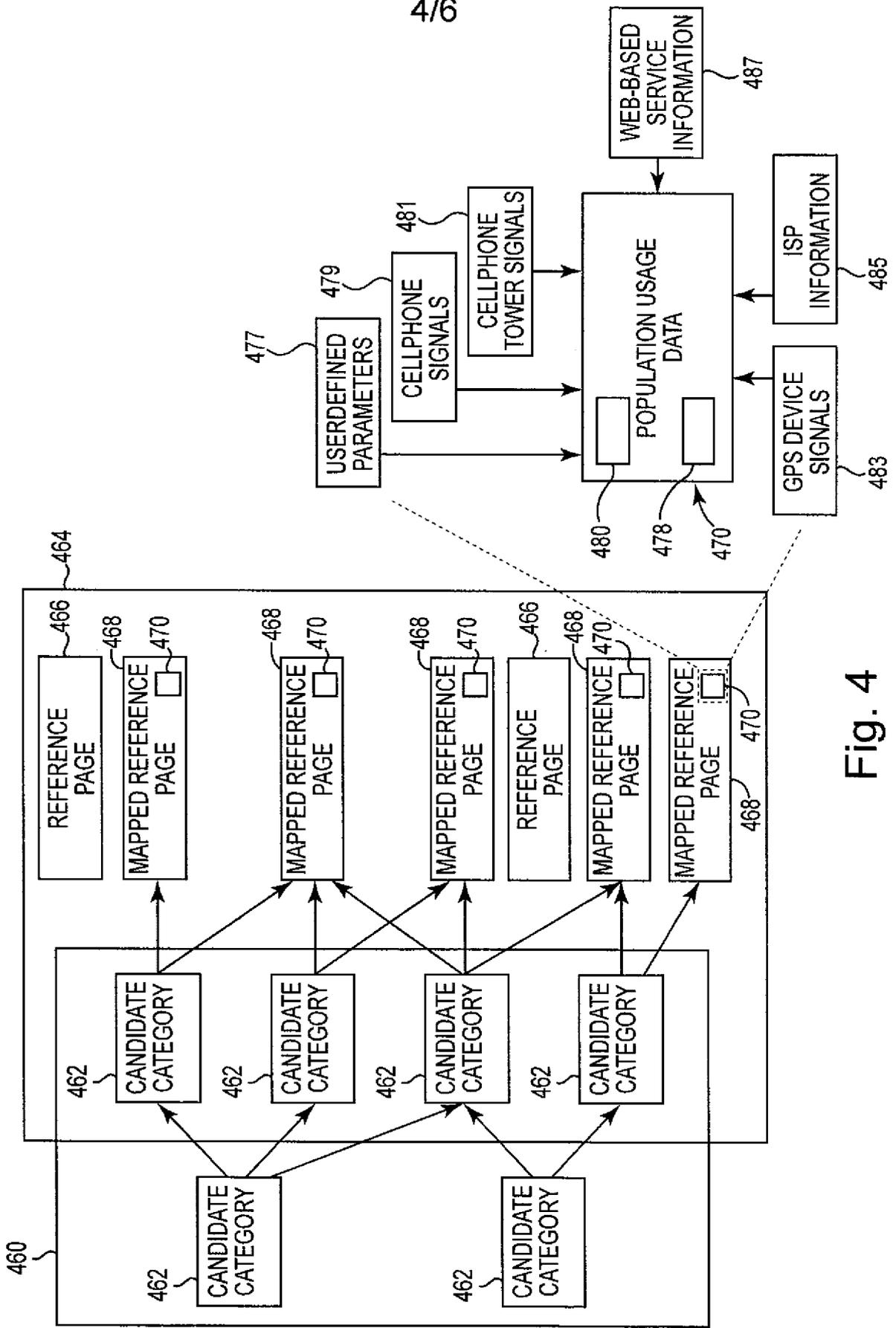


Fig. 4

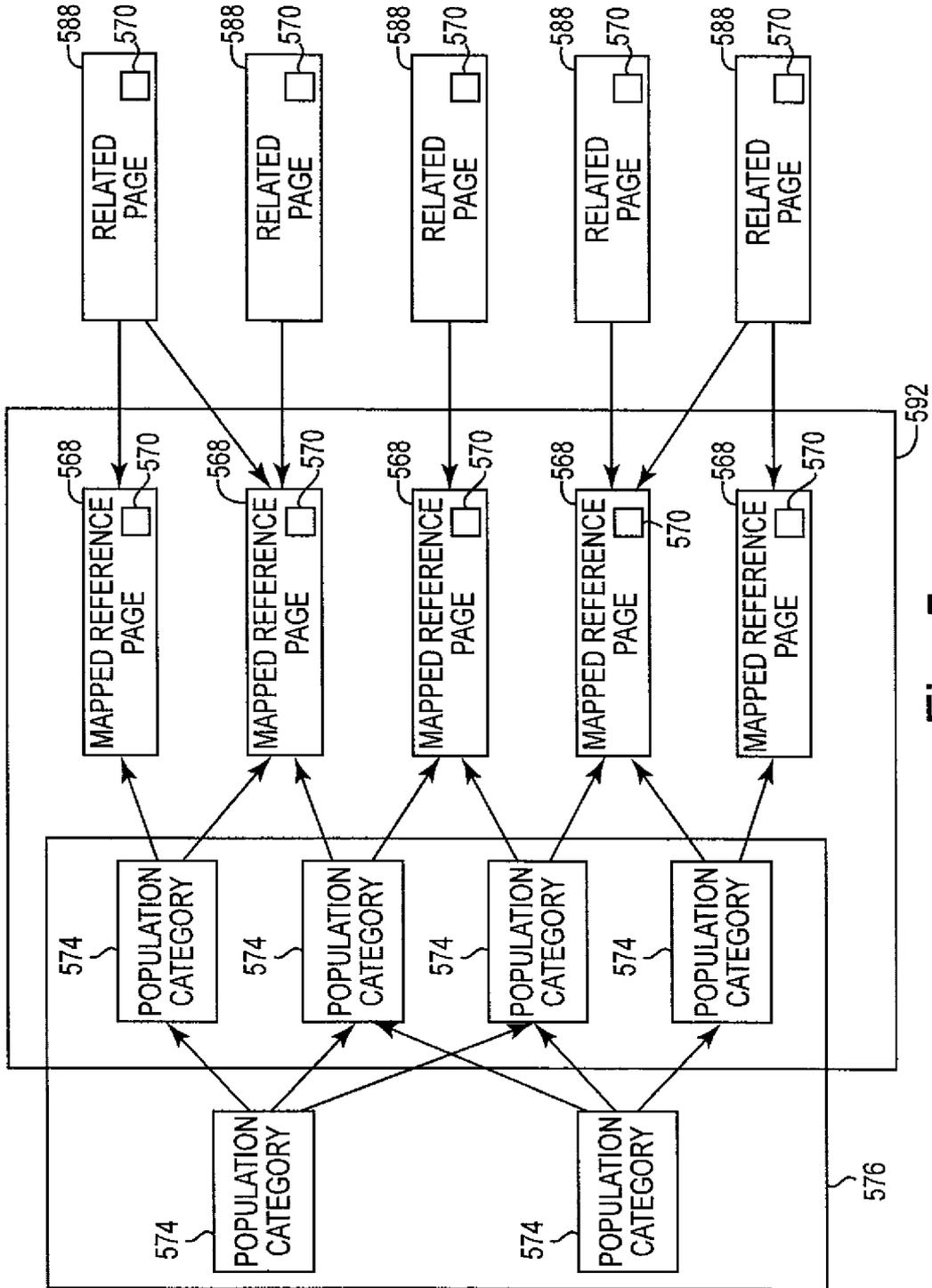


Fig. 5

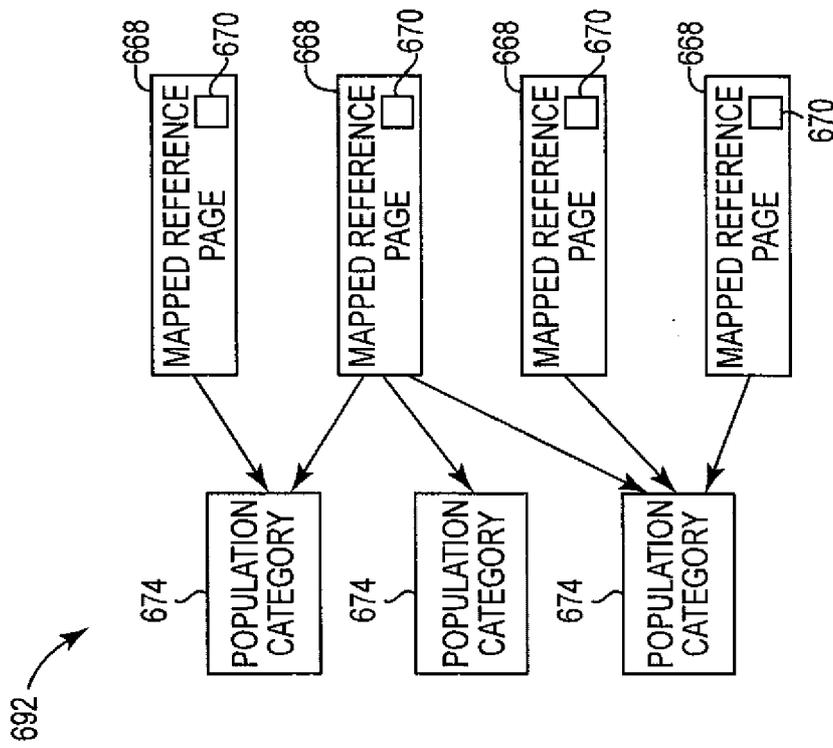


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2011/025030**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/00(2006.01)i, G06F 17/30(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)

G06F 17/00; G06F 15/173; G06F 7/00; G06F 17/30

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords:web, category, hierarchy, mapping

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2001-0047356 AI (TOM REN et al.) 29 November 2001 See Abstract ; paragraph 17; paragraph 19; paragraph 21; figures 1-2 .	1-15
A	US 2007-0038655 AI (BRONSTAD P. MATTHEW et al.) 15 February 2007 See Abstract ; paragraph 34 - paragraph 38; figures 1-2	1-15
A	US 2008-0109420 AI (BRITTON COLIN P. et al.) 08 May 2008 See paragraph 14; paragraph 78;	1-15
A	US 2008-0263200 AI (OR SIM LIM et al.) 23 October 2008 See paragraph 59; paragraph 63	1-15

 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 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

27 SEPTEMBER 2011 (27.09.2011)

Date of mailing of the international search report

28 SEPTEMBER 2011 (28.09.2011)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Cheongsu-ro,
Seo-gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

PARK, SANG HYUN

Telephone No. 82-42-481-8263



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US201 1/025030

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2001-0047356 A1	29.11.2001	None	
US 2007-0038655 A1	15.02.2007	US 2008-020856 1 A1 us 7725407 B2	28.08.2008 25.05.2010
US 2008-0 109420 A1	08.05.2008	AU 2002-314788 A8 AU 2002-3465 10 A1 AU 2002-3465 10 A8 AU 2002-365577 A1 AU 2003-234177 A1 AU 2003-279845 A1 AU 2003-279845 A8 CA 2447322 A1 CA 247 1467 A1 CA 247 1468 A1 CA 2483994 A1 CA 250 1847 A1 EP 1399838 A1 EP 1399838 A4 EP 14052 19 A2 EP 14052 19 A4 EP 1483688 A1 EP 1502253 A1 EP 1502253 A4 EP 1546921 A2 EP 1588286 A2 EP 16902 10 A2 US 2002-0174126 A1 US 2002-0178 170 A1 US 2003-0158841 A1 US 2003-015885 1 A1 US 2003-0208499 A1 US 2004-0073545 A1 US 2005-0055330 A1 US 2005-0187926 A1 US 2005-0228805 A1 US 2006-0036620 A1 US 2006-027 1563 A1 US 2006-0277227 A1 us 2007-0198454 A1 us 2008-0109485 A1 us 6856992 B2 us 6925457 B2 us 6954749 B2 us 7058637 B2 us 7302440 B2 us 73 18055 B2 us 76 137 12 B2 us 7640239 B2	25.11.2002 10.06.2003 10.06.2003 10.06.2003 17.11.2003 04.05.2004 04.05.2004 21.11.2002 30.05.2003 05.06.2003 13.11.2003 22.04.2004 24.03.2004 21.06.2006 07.04.2004 28.06.2006 08.12.2004 02.02.2005 19.09.2007 29.06.2005 26.10.2005 16.08.2006 21.11.2002 28.11.2002 21.08.2003 21.08.2003 06.11.2003 15.04.2004 10.03.2005 25.08.2005 13.10.2005 16.02.2006 30.11.2006 07.12.2006 23.08.2007 08.05.2008 15.02.2005 02.08.2005 11.10.2005 06.06.2006 27.11.2007 08.01.2008 03.11.2009 29.12.2009

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2011/025030

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		US 783 1604 B2	09 . 11 . 2010
		US 78905 17 B2	15 . 02 . 201 1
		wo 02-0933 19 A2	2 1 . 11 . 2002
		wo 02-0933 19 A3	2 1 . 11 . 2002
		wo 02-093410 A1	2 1 . 11 . 2002
		wo 03-044634 A2	30 . 05 . 2003
		wo 03-044634 A3	30 . 05 . 2003
		wo 03-046769 A1	05 . 06 . 2003
		wo 03-094142 A1	13 . 11 . 2003
		wo 2004-034625 A2	22 . 04 . 2004
		wo 2004-034625 A3	22 . 04 . 2004
		wo 2005-029365 A2	3 1 . 03 . 2005
		wo 2005-029365 A3	3 1 . 03 . 2005
US 2008-0263200 A1	23 . 10 . 2008	AU 2000-62543 A1	05 . 03 . 2001
		AU 2000-62543 B2	28 . 04 . 2005
		AU 6254300 A	05 . 03 . 2001
		AU 78 1008 B2	28 . 04 . 2005
		EP 1226525 A1	3 1 . 07 . 2002
		EP 1226525 A4	09 . 08 . 2006
		EP 2293495 A2	09 . 03 . 201 1
		JP 2003-506790 A	18 . 02 . 2003
		JP 2011-018384 A	27 . 0 1 . 201 1
		NZ517638A	3 1 . 10 . 2003
		US 2008-0256235 A1	16 . 10 . 2008
		US 2010-0228855 A1	09 . 09 . 2010
		US 2011-0208860 A1	25 . 08 . 201 1
		US 7376722 B1	20 . 05 . 2008
		US 7953791 B2	3 1 . 05 . 201 1
		US 7953839 B2	3 1 . 05 . 201 1
		wo 01-11506 A1	15 . 02 . 2001
		wo 01-11506A1	15 . 02 . 2001