

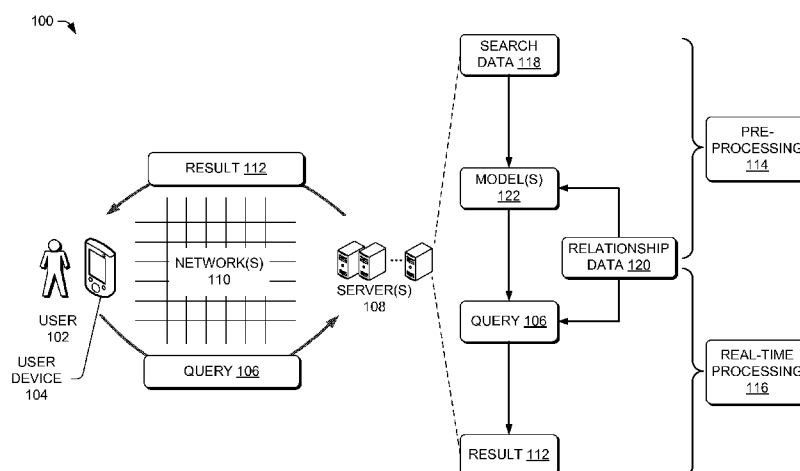


- (51) **International Patent Classification:**
G06F 17/30 (2006.01)
- (21) **International Application Number:**
PCT/US2012/065096
- (22) **International Filing Date:**
14 November 2012 (14.11.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
201110361975.3 15 November 2011 (15.11.2011) CN
- (71) **Applicant:** ALIBABA GROUP HOLDING LIMITED
[—/US]; Fourth Floor, One Capital Place, P.O. Box 847,
Grand Cayman (KY).
- (72) **Inventors:** LANG, Hao; 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). TANG, Chao; 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). ZHANG, Xiaoxun; 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). XUE, Guirong; 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) **Agents:** NELSON, Brett, L. et al.; 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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **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, 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))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) **Title:** SEARCH METHOD, SEARCH APPARATUS AND SEARCH ENGINE SYSTEM

(57) **Abstract:** Techniques for performing searches using models generated from data mining. These techniques may include receiving a query including a search string. If a portion of the search string matches with a category of relationship data, the portion of the search string may be replaced with a label of the category to generate a modified query. Based on the modified query and models generated from mining historical search data, intent of the query may be determined, and searches may be then performed.

Search Method, Search Apparatus and Search Engine System

CROSS REFERENCE TO RELATED PATENT APPLICATIONS

This application claims priority to Chinese Patent Application No.
5 201110361975.3, filed on November 15, 2011, entitled "Search Method,
Search Apparatus and Search Engine System," which is hereby incorporated by
reference in its entirety.

TECHNICAL FIELD

10 This disclosure relates to the field of network technologies. More
specifically, the disclosure relates to a search method, a search device and a
search engine system.

BACKGROUND

15 Search engine systems gather information from the Internet and
present users with search results in response to queries. Conventional search
engine systems perform searches based on the results' popularity and
relevancy. However, sometime the results do not match the user intent of
searches. For example, a user may intend to download a song "Firefly" sung
20 by a singer "Wang Fei." But conventional search engine systems may return
links associated with insects as top ranks. To avoid this situation, many
different approaches have been proposed (e.g., category searches). While

important progress has been made, the existing approaches present problems, e.g., low accuracy and efficiency for Internet searching.

SUMMARY

5 Described herein are techniques for performing searches using models generated from data mining. The techniques may analyze historical search data to generate models including intended words and labels of categories of relationship data. These models may be used to analyze queries to determine intent associated with the queries. Based on the queries and intent, the
10 techniques may then perform searches.

This Summary is not intended to identify all key features or essential features of the claimed subject matter, nor is it intended to be used alone as an aid in determining the scope of the claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

15 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 illustrates an example architecture that includes server(s) for
20 performing data mining and searches.

FIG. 2 illustrates an example flow diagram for data mining.

FIG. 3 illustrates an example flow diagram for performing searches.

FIG. 4 illustrates another example flow diagram for performing searches.

FIG. 5 illustrates an example user interface served by the server(s) of FIG. 1.

FIG. 6 illustrates an example server that may be deployed in the
5 architecture of FIG. 1

DETAILED DESCRIPTION

The discussion below describes specific exemplary embodiments of the present disclosure. The exemplary embodiments described here are for
10 exemplary purposes only, and are not intended to limit the present disclosure.

FIG. 1 illustrates an example architecture 100 that includes server(s) for perform data mining and searches. A user may submit a query to a server, and the server may perform searches and return results. In some embodiments, the server may conduct data pre-processing (e.g., offline data processing)
15 before performing searches. The data pre-processing may be conducted on historical search data including a large scale of queries received by a search engine system. Based on the historical data, the server may generate models that indicate user intents and categories associated with the queries. In some embodiments, the servers may modify the query to match a model
20 corresponding to the query, and then performed a search based on the model and the query.

In the illustrated embodiment, the techniques are described in the context of a user 102 operating a user device 104 to submit a query 106 to one or more server(s) 108 over one or more network(s) 110. For instance, the user 102 may submit the query 106 including terms of "Firefly MP3 download."

- 5 The server 108 may perform a search based on these terms, and return a result 112 to the user device 104.

Here, the user 102 may submit the query 106 via network 110. The network 110 may include any one or combination of multiple different types of networks, such as cable networks, the Internet, and wireless networks. The user device 104, meanwhile, may be implemented as any number of computing devices, including as a personal computer, a laptop computer, a portable digital assistant (PDA), a mobile phone, a set-top box, a game console, a personal media player (PMP), and so forth. The user device 104 is equipped with one or more processors and memory to store applications and data. An application, such as a browser or other client application, running on the user device 104 may facilitate submission to the server 108 over network 110.

In architecture 100, the server 108 may perform a pre-processing 114, and then perform a real-time processing 116 in response to the query 106. The server 108 may perform the pre-processing 114 using search data 118 and relationship data 120. The search data 118 may include big-scale historical search data (e.g., millions queries and their results), and the relationship data 120 may include relational data structures of specific fields (e.g. song titles and

artist names corresponding to the song titles). The server 108 may apply data mining technologies to generate models 122 indicating user intents associated with queries. A user intent may be indicated by one or more particular intent words/characters. For example, in the field of music, intended words may be

5 “mp3 download” and “free download,” and the model may be indicated as: “<Song>mp3 download,” “<Song>free download” or “<Song> <Singer>.” Here, <Song> and <Singer> represent the title of a song and a name of an artist, respectively. The pre-processing 114 is described in a greater detail in FIG. 2.

In the illustrated embodiment, the server 108 may perform the real-time

10 processing 116 based on the models 122, the relationship data 120 and the query 106 transmitted by the user device 104, which is described in a greater details in FIGS. 3 and 4.

FIG. 2 illustrates an example flow diagram 200 for data mining to generate search models. In some embodiments, the search data 118 may

15 include log data in a search engine system. The log data may record historical search strings submitted by users. The historical search strings may be one or more keywords, phrases or sentences. The relationship data 120 may include data in relational data structures used by specific fields (e.g., news, images, videos, maps, music, or shopping). As for search data of each field, the

20 relational data structures can be used to store them. In some embodiments, the relational data 120 may be represented using relational math model (e.g.,

a 2-dimensional table). For example, in the field of music, a song title and the artist's name make up a pair of relationship data.

In the illustrated embodiment, at 202, the server 108 may collect the search data 118 that includes multiple queries. The multiple queries may
5 include multiple historical search strings of specific fields.

At 204, the server 108 may modify the multiple queries. The server 108 may determine whether individual historical search string matches with a category of the relationship data of a specific field in a predetermined order. For example, the category may be indicated by metadata in the relationship
10 data of the specific field. If a portion of the historical search string matches the metadata of any category, the server 108 may replace the portion with a label of the metadata category.

For example, the historical search string includes "Firefly Wang Fei MP3 download." If a match is found for song title "Firefly" and artist name "Wang
15 Fei," the server 108 may change the string to "<Song> <Singer> MP3 download," wherein <Song> and <Singer> correspond to the labels of the song title and artist name, respectively.

At 206, the server 108 may rank the query based on a frequency of the modified query in the search data 118. In some embodiments, the server 108
20 may calculate a number of times/frequency that the historical search string appears, and then sort multiple queries in the search data 118 based on a frequency associated with individual query.

The server 108 may also determine historical search strings having frequencies being greater than a predetermined number. Based on these historical search strings, the server 108 may generate the models 122 at 208. In some embodiments, the matched historical search string that exceeds the
5 frequency threshold may be used as the models 122. Accordingly, words/characters in the model may be used to indicate user intents associated with the multiple queries via intended words/characters.

For example, models for the field of music may include "<Song> <Singer>," "<Song> mp3 download," "<Song> ringtone <Singer>" and etc. In
10 some embodiments, the model may include at least one of a label of category of the relationship data 120 and intended words obtained from the pre-processing 114. In these instances, a stand-alone intended word cannot make up a model. Also, the word that appears in a model may be the intended words, while a word that does not appear in a model may not be the intended
15 words.

FIG. 3 illustrates an example flow diagram 300 for performing searches using models generated during data mining. The server 108 may perform an online matching process using models that are generated based on the search data 118. In some embodiments, the server 108 may combine generated
20 intended words and labels of metadata categories of the relationship data 120 to generate data (e.g., a dictionary), which may be used for the online matching process.

At 302, the server 108 may receive a query from the user device 104. The query may include a string of words, characters, and/or symbols. At 304, the server 108 may perform online matches based on the received query. In some embodiments, a portion of the query may be compared against the dictionary, from left to right. In these instances, segmentation of the string may not be performed. For example, a single word as a unit may be used during the online match. If the string includes ideograms (e.g., Chinese), an individual character may be used as the unit. If the string includes alphabet (e.g., English), each word by means of the spaces/blanks may be distinguished. In these instances, an individual word may be used as the unit. For conventional technologies, searches are performed after the string has been segmented. However, segmentation is a slow process, and strings including new words may be incorrectly segmented, therefore causing inaccuracies for searches performed thereafter.

If a portion of the search string matches the metadata in the relationship data, the server 108 may replace the portion with a label of a corresponding metadata category to generate a modified search string. If a portion of the search string does not match the dictionary (i.e., metadata categories in the dictionary) and does not match intended words included in the dictionary, the server 108 may terminate the matching process. If the search string matches the dictionary from left to right and the modified search string matches a model generated during the pre-processing 114, the server 108 at 306 may

determine search intent associated with the search string. At 308, the server 108 may perform searches based on the search intent.

In some embodiments, the search system may consider the user's level of attention in categorizing the relationship data 120 of a particular field (e.g., music, novel, film and other categories). The server 108 may then combine intended words of different fields to generate different types of dictionaries. During the search, the server 108 may first consider dictionaries of popular fields, and then non-popular fields. In some embodiments, the non-popular fields may not be considered unless there is no match in the popular fields.

Suppose that the dictionaries of popular fields include three fields: music, novel and film. After the user device 104 submits a search string: "MP3 Firefly Wang Fei", the server 108 may not perform segmentation, but analyze meaning of words or spaces/blanks. The server 108 may then recognize individual word, and separately perform matching with the dictionaries of the popular fields.

During the matching process, the server 108 may separately match the word "MP3" with the intended words and metadata in the dictionaries of music, novel, and film. Further suppose that the word "MP3" is found in the dictionaries of music and film. As a result, MP3 may be regarded as an intended word, and therefore no replacement may be performed. Then, the server 108 may match the character "Ying," then the character "Huo" and finally the character "chong" in the dictionaries of music and film.

Further suppose that all three characters are found in the dictionaries of music and film. As a result, the server 108 may replace the word “Firefly” (i.e., Ying Huo Chong) with a label <Song> in a music model based on the current models. In a film model, the server 108 may replace the word “Firefly” (i.e.,
5 Ying Huo Chong) with a label <Film>. After replacement, the match may continue to be performed to “Wang” and “Fei.” Similarly, the words “Wang Fei” may be replaced with the label <Singer>. Accordingly, the model obtained after matching and replacement is “MP3 <Song> <Singer>”, and this model may exist among the current models. Therefore, the server 108 may
10 determine that the user’s search intent is to find a song “Firefly” that is sung by Wang Fei.

In the above online matching process, although the word “Firefly” may also be found in other dictionaries (e.g., an insect category), the server 108 may not consider a category of insects since the category is not popular
15 compared to the previously discussed categories of film, music and novels.

FIG. 4 illustrates another example flow diagram 400 for performing searches. At 402, the server 108 may receive a query including a search string. At 404, the server 108 may determine whether a portion of the search string matches with a category of pre-generated dictionary. In some embodiment,
20 the dictionary may be a combination of intended words/characters that are generated from the pre-processing 114, and labels of categories of the relationship data 120.

In some embodiments, the matching may be performed between the search string, starting from left to right, and the pre-generated dictionary. In the matching process, there is no need to perform segmentation on the real-time search string, but an individual word may be used as the unit in performing the match, successively. The matching may be total matching including matches between all contents of the search string and intended words or labels of metadata in the dictionary. In some embodiments, since the elements in the generated models may have orders (e.g., "<Song> <Singer>" and "<Singer> <Song>"), the matching may be assigned a predetermined order (e.g., from left to right).

If a portion of the search string matches a category of categories in the dictionary, the server 108 may replace the portion with a label of the category at 406. After the replacement, the server may at 408 generate the modified search string that may include at least one of a label of the category, intended words and the unmatched portion. For example, if the search string "Firefly Wang Fei MP3 download" matches the song title "Firefly" and artist name "Wang Fei" in the metadata category, the server 108 may replace the search string with the labels to generate a modified search string: "<Song> <Singer> MP3 download".

At 410, the server 108 may determine search intent associated with the search string. At 412, the server 108 may perform searches based on the search string and the search intent. If the matching is performed successfully,

the modified search string may include at least one of a label of the metadata category of the relationship data 120, a label of another category that include intended words, and unmatched portions of the search string.

For example, if the search string "Firefly Wang Fei MP3 download" matches the song title "Firefly" and artist name "Wang Fei" in metadata categories of the relationship data 120, the search string may be modified to include the intended words "MP3 download" and labels of the metadata categories. Accordingly, the search string "Firefly Wang Fei MP3 download" matches with a model "<Song> <Singer> MP3 download." This may help recognize that the search intent of the search string is to "download the song Firefly that was sung by Wang Fei". On the other hand, if the model "<Song> <Singer> free" does not exist among the pre-generated models, the search string and the model may be deemed as mismatching.

In some embodiments, the server 108 may place results data that correspond to a search intent on the first spot of returned search results. Also, the results can include several related metadata information. If the user 102 clicks the link to the metadata information, the server may lead the user to a webpage containing details associated with the metadata information.

For example, the results may include a song "Firefly" sung by "Wang Fei." The server 108 may place some related information on the first spot of the returned results. The related information may include the song title, artist name, album name, lyrics, file size etc. In these instances, each piece of

information may be metadata, and the structure of the metadata is connected/inter-related. In addition, each piece of the metadata can be set up as a link.

FIG. 5 illustrates an example user interface (UI) 500 served by the server(s) of FIG. 1. The UI 500 may be presented by the user device 104 or the server 108. The UI 500 may include a result section 502 including multiple results (e.g., a result 504 and a result 506). The result 504 may include multiple sections for detailed information regarding the search result. In the illustrated embodiments, the result 504 may include multiple subsections, e.g., a song title 508, a singer name 510, an album title 512 the song belongs to, a music format 514, a lyric source 516 and a download link 518.

In some embodiments, the multiple subsections may be expanded for the user 102 to review in more detail. In the illustrated embodiments, the singer name 510 may be expanded to include multiple song subsections indicating songs presented by the singer (e.g., subsection 520 and subsection 522). The expanded download link 518 may include a reminder 524 that indicates the song is being downloaded. The expanded download link 518 may also include a retry link 526 to enable the user 102 to retry if a previous download is not successful.

FIG. 6 illustrates an example server 108 that may be deployed in the architecture of FIG. 1. The server 108 may be configured as any suitable computing device(s). In one exemplary configuration, the server 108 includes

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
5 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,
10 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
15 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
20 include transitory media such as modulated data signals and carrier waves.

Turning to the memory 608 in more detail, the memory 608 may include an input module 610, a real-time matching module 612, a replacement

module 614, a real-time output module 616 and a model generation module 618.

The input module 610 may be configured to receive a real-time search string. The real-time matching module 612 may be configured to perform
5 matching between the real-time search string and the pre-generated dictionary, wherein the the dictionary is a combination of the first category word and the second category word. The replacement module 614 may be configured to replace a portion with a label of the category where the second category word belongs when the portion of the real-time search string
10 matches the second category word of any category in the dictionary. After the replacement, the real-time search string may include the label of the category where the first category word and/or the second category word belongs, and/or the unmatched portion.

The real-time output module 616 may be configured to determine the
15 specific search intent of the matched real-time search string, and to return the search result data based on the specific search intent based on the pre-generated models. The pre-generated models may include at least one of metadata category labels and intended words. Correspondingly, the first category word in the dictionary may be the intended word in the models, and
20 the second category word in the dictionary can be the metadata in the relationship data.

In some embodiments, if the matched real-time search string includes the label of the category where the second category word belongs, and/or, is made up of the label of the category where the first category word and second category word belong, then the matched real-time search string and any of
5 the pre-generated models may be matched with each other. The matched real-time search string may include the specific search intent stated in the model.

The model generation module 618 may include a replacement sub-module configured to match the historical search string in the historical search
10 data with the metadata in the relationship data in a predetermined order. If a portion of the historical search string matches the metadata of any category, the portion may be replaced with the label of the category of the metadata.

The model generation module 618 may also include a calculation sub-module configured to calculate the number of times/frequency that the
15 matched historical search string appears, and perform sorting based on the calculated frequency. Then the model generation module 618 may use the matched historical search string that exceeds the frequency threshold as the model.

The present disclosure discloses an exemplary embodiment for a type of
20 search engine system. In the exemplary embodiment, the search engine system may include the server 108 mentioned in any of the above exemplary embodiments. The exemplary embodiments of this disclosure may be

implemented by a search device (e.g., the server 108) and/or search engine system, which may be used in the field of searching, especially web searching and mobile searching.

The present disclosure can be described in the context of computer
5 executable instructions, such as a program module. In general, a program module includes executing specific tasks or implementing specific abstract data type routines, programs, objects, components, data structures, etc. The present disclosure can also be implemented in a distributed computing environment. In these distributed computing environment, tasks are executed
10 using remote processing equipment, which are connected using communication networks. In a distributed computing environment, the program module can be stored in the local and remote computer storage media.

Here, relationship terms such as “first” and “second” are solely used to
15 separate one entity or operation from another entity or operation, and not necessarily to require or imply that such actual relationships or sequences exist among these entities or operations. Also, the “and/or” indicates that this document comprises the “and” relationship, and also comprises the “or” relationship, wherein: if object A and object B have an “and” relationship, then
20 it means that an exemplary embodiment can include both object A and object B at the same time; if object A and object B have an “or” relationship, then it

means that an exemplary embodiment can include only object A, or only object B.

The specific examples herein are utilized to illustrate the principles and embodiments of the application. The description of the embodiments above is
5 designed to assist in understanding the method and ideas of the present disclosure. However, persons skilled in the art could, based on the ideas in the application, make alterations to the specific embodiments and application scope, and thus the content of the present specification should not be construed as placing limitations on the present application.

CLAIMS

What is claimed is:

1. One or more computer-readable media storing computer-executable instructions that, when executed by one or more processors, instruct the one

5 or more processors to perform acts comprising:

receiving a query including a string;

determining that a portion of the string matches with a category of relationship data;

10 replacing the portion of the string with a label of the category to generate a modified query;

analyzing the modified query to determine an intent of the query based on search models; and

performing a search based on the intent.

15 2. The one or more computer-readable media of claim 1, wherein the analyzing the modified query to determine the intent of the query based on the search models comprises analyzing the modified query to determine the intent of the query using a model of the search models that matches with the modified query.

20

3. The one or more computer-readable media of claim 1, further comprising identifying an intention indicator included in the string, wherein the modified query includes the intention indicator and the label.

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

collecting multiple queries from a historical search data;

determining that a portion of individual query of the multiple queries matches with one category of the relationship data;

10 modifying the individual query by replacing the portion of the individual query with one label of the one category of the relationship data to generate modified individual query;

calculating a frequency that the modified individual query appears in the historical search data; and

15 generating the search models using modified individual query based on the frequency.

5. The one or more computer-readable media of claim 4, wherein the generating the models using modified individual query based on the frequency

20 comprise:

ranking the multiple queries in the historical search data based on the frequency; and

generating the search models using queries of the multiple queries that have frequencies greater than a predetermined number.

6. The one or more computer-readable media of claim 1, wherein the
5 models are generated by analyzing historical search data and the relationship data.

7. The one or more computer-readable media of claim 6, wherein the
historical search data is determined based on log data generated by a search
10 engine system.

8. The one or more computer-readable media of claim 1, wherein the
string includes at least one of a word, a character, a symbol, a phrase and a
sentence.

15 9. The one or more computer-readable media of claim 1, wherein the
relationship data includes a relational data structure in a particular field.

10. The one or more computer-readable media of claim 9, wherein the
20 particular field is at least one of news, an image, a map and music.

11. The one or more computer-readable media of claim 1, wherein the determining comprises determining without conducting segmentation of the string.

12. A computer-implemented method comprising:

5 receiving a query including search string;

determining that a portion of the search string matches with a category of relationship data;

identifying an intention indicator included in the search string;

10 generating a modified search string including a label of the category of the relationship data and the intention indicator;

determining an intent of the query based on the modified search string and models; and

performing a search based on the determined intent.

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

collecting multiple queries from a historical search data;

determining that a portion of individual query of the multiple queries matches with one category of the relationship data;

modifying the individual query;

20 calculating a frequency that the modified individual query appears in the historical search data; and

generating the models using modified individual query based on the frequency.

14. The computer-implemented method of claim 13, wherein the
5 modifying the individual query comprises modifying the individual query by replacing the portion of the individual query with one label of the category of the relationship data to generate modified individual query.

15. The computer-implemented method of claim 13, wherein the
10 generating the models using modified individual query based on the frequency comprise:

ranking the multiple queries in the historical search data based on the frequency; and

generating the models using queries of the multiple queries that have
15 frequencies greater than a predetermined number.

16. The computer-implemented method of claim 12, wherein the relationship data includes a relational data structure in a particular field.

20 17. The computer-implemented method of claim 16, wherein the particular field is at least one of news, an image, a map and music.

18. A computing device comprising:

one or more processors; and

memory to maintain a plurality of components executable by the one or more processors, the plurality of components comprising:

- 5 an input module that receives a query including a string,
 a real-time matching module that determines that a portion of
the string matches with a category of relationship data,
 a replacement module that replaces the portion of the string
with a label of the category to generate a modified query;
10 a real-time output module that:
 determines an intent of the query based on models that
are generated by analyzing historical search data and the
relationship data, and
 performs a search based on the query and the
15 determined intent.

19. The computing device of claim 18, further comprising a model
generation module that:

collects multiple queries from a historical search data;

- 20 determines that a portion of individual query of the multiple queries
matches with one category of the relationship data;

modifies the individual query by replacing the portion with one label of the one category of the relationship data to generate modified individual query;

calculates a frequency that the modified individual query appears in the
5 historical search data;

ranks the multiple queries in the historical search data based on the frequency; and

generates the models using queries of the multiple queries that have frequencies greater than a predetermined number.

10

20. The computing device of claim 18, wherein the string includes at least one of a word, a character, a symbol, a phrase and a sentence.

100 →

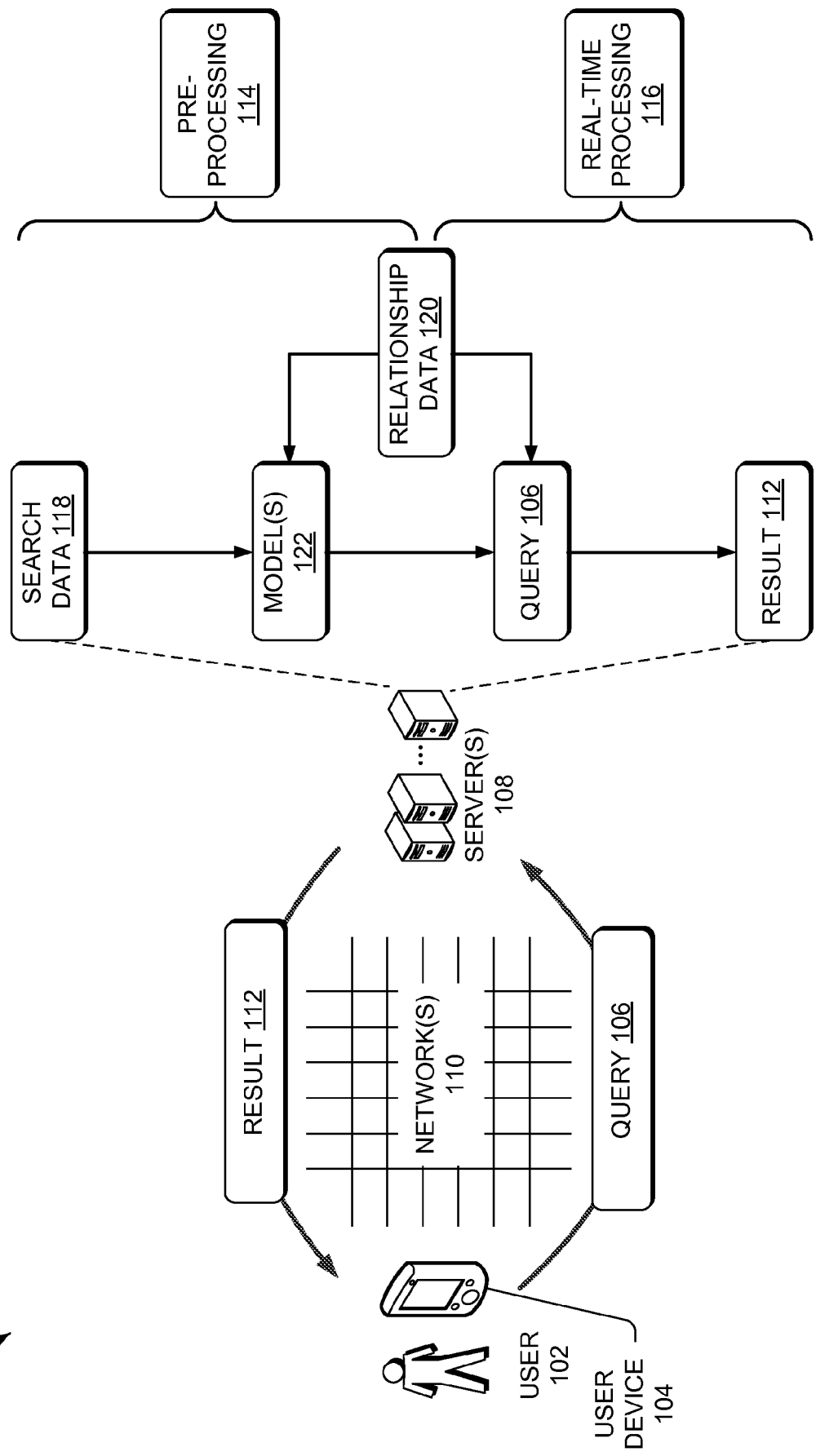


FIG. 1

200

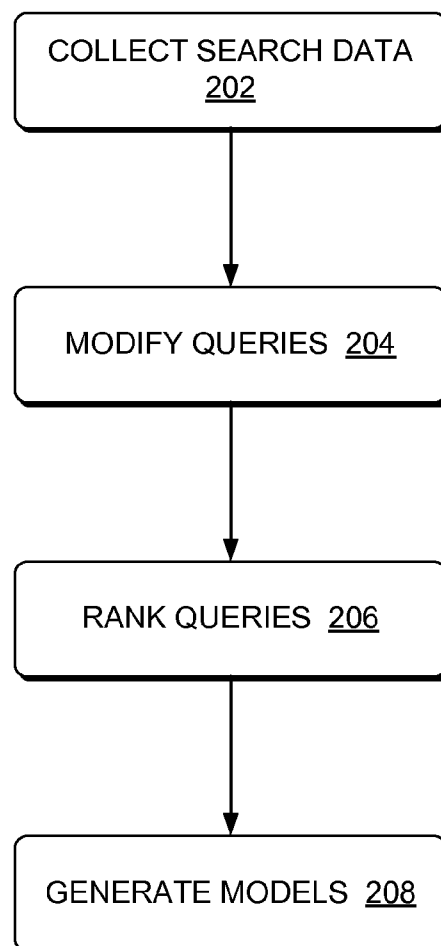


FIG. 2

↖ 300

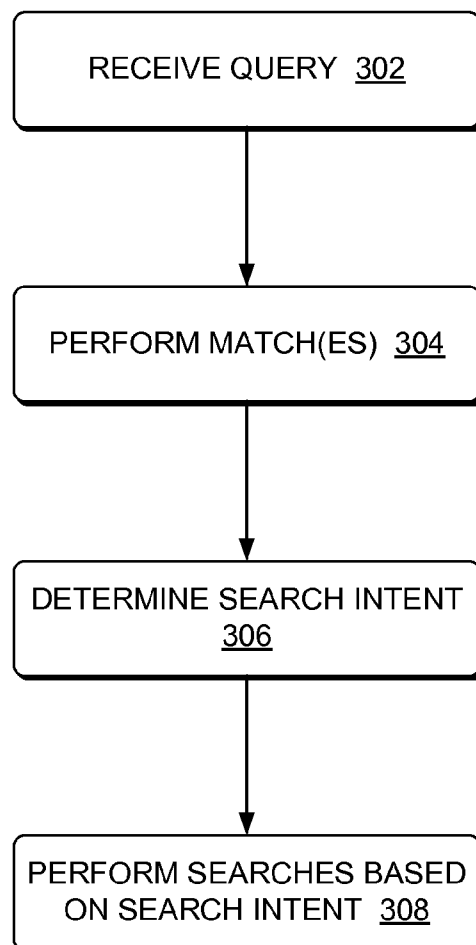


FIG. 3

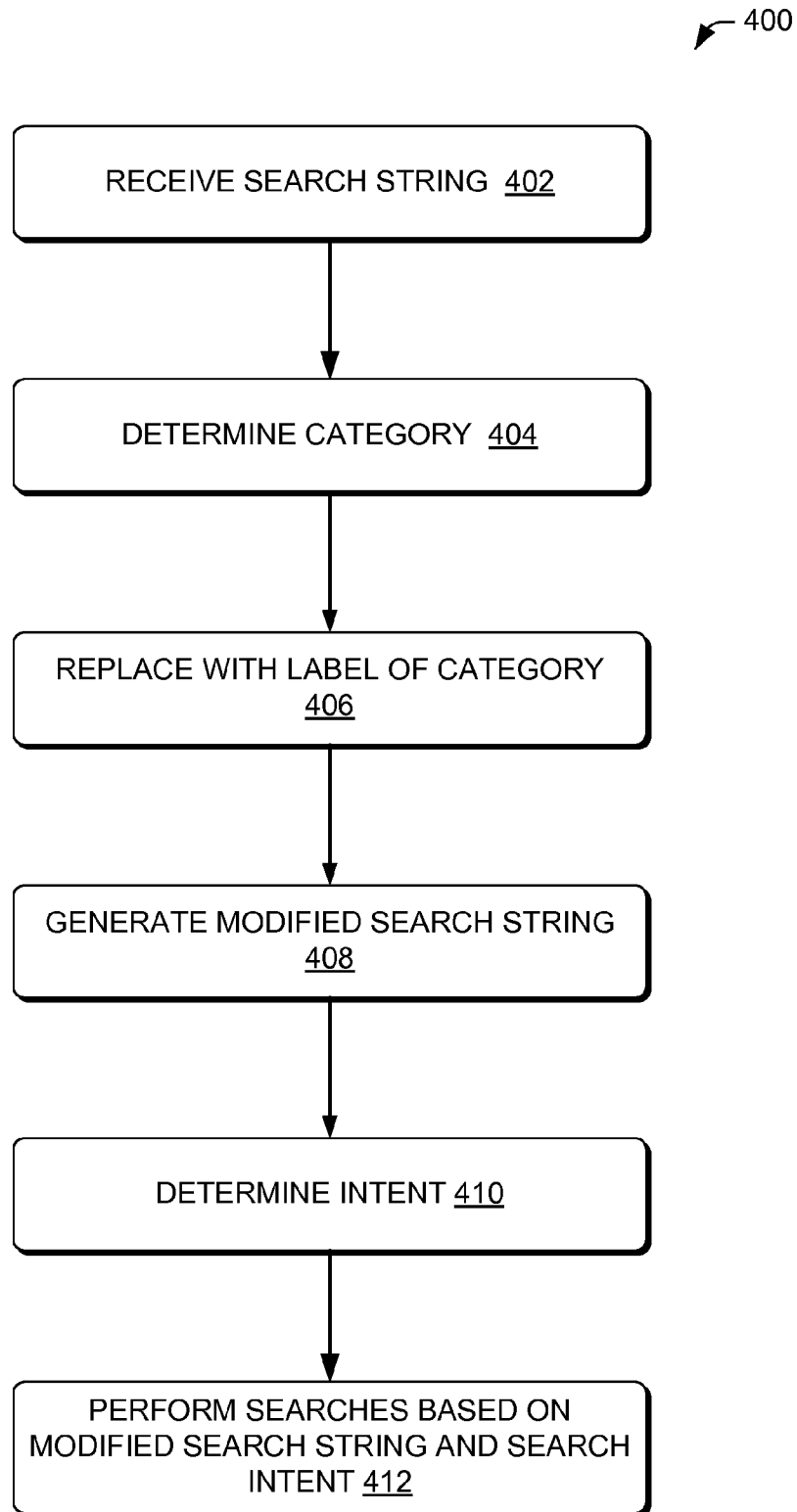


FIG. 4

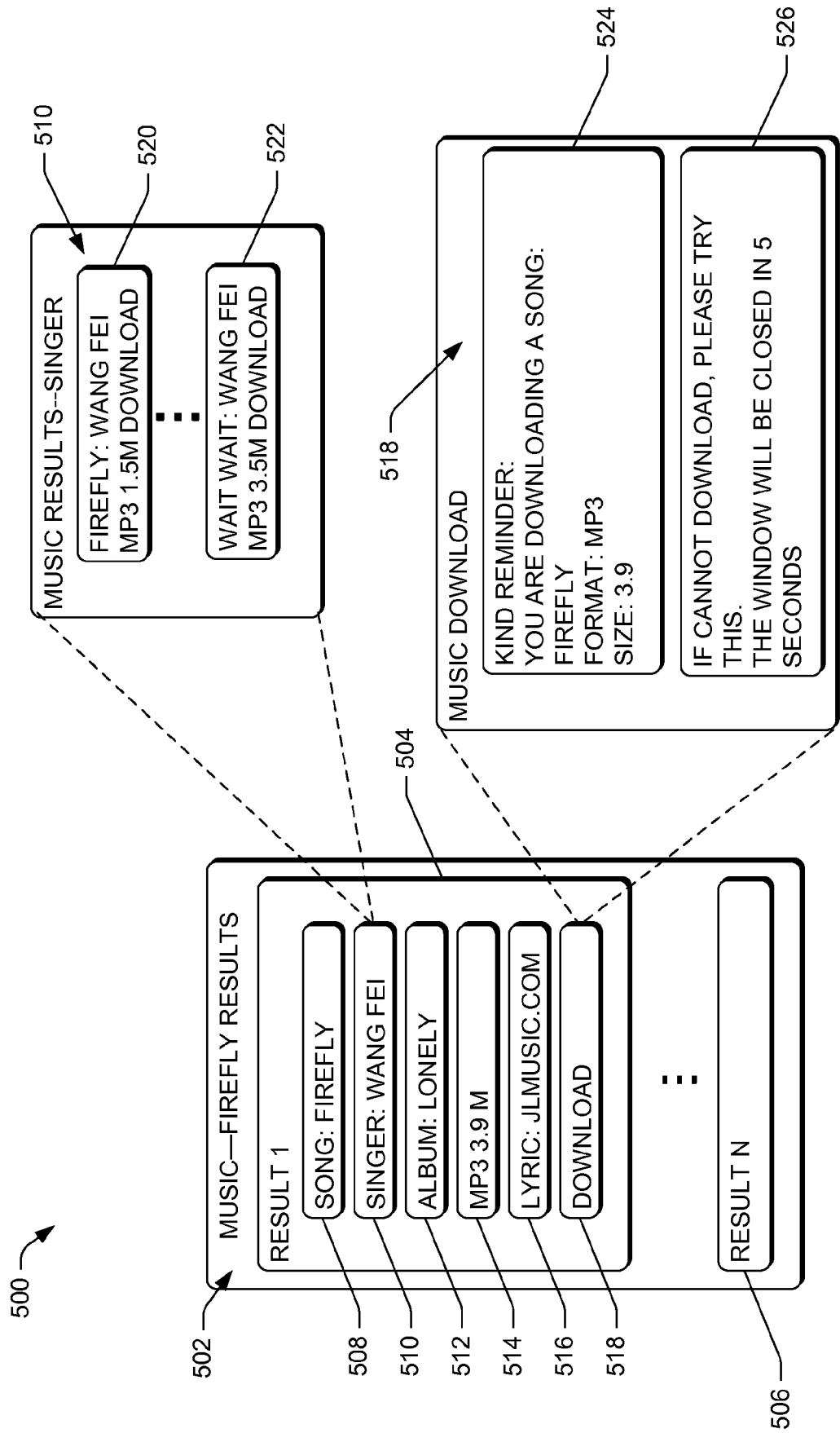


FIG. 5

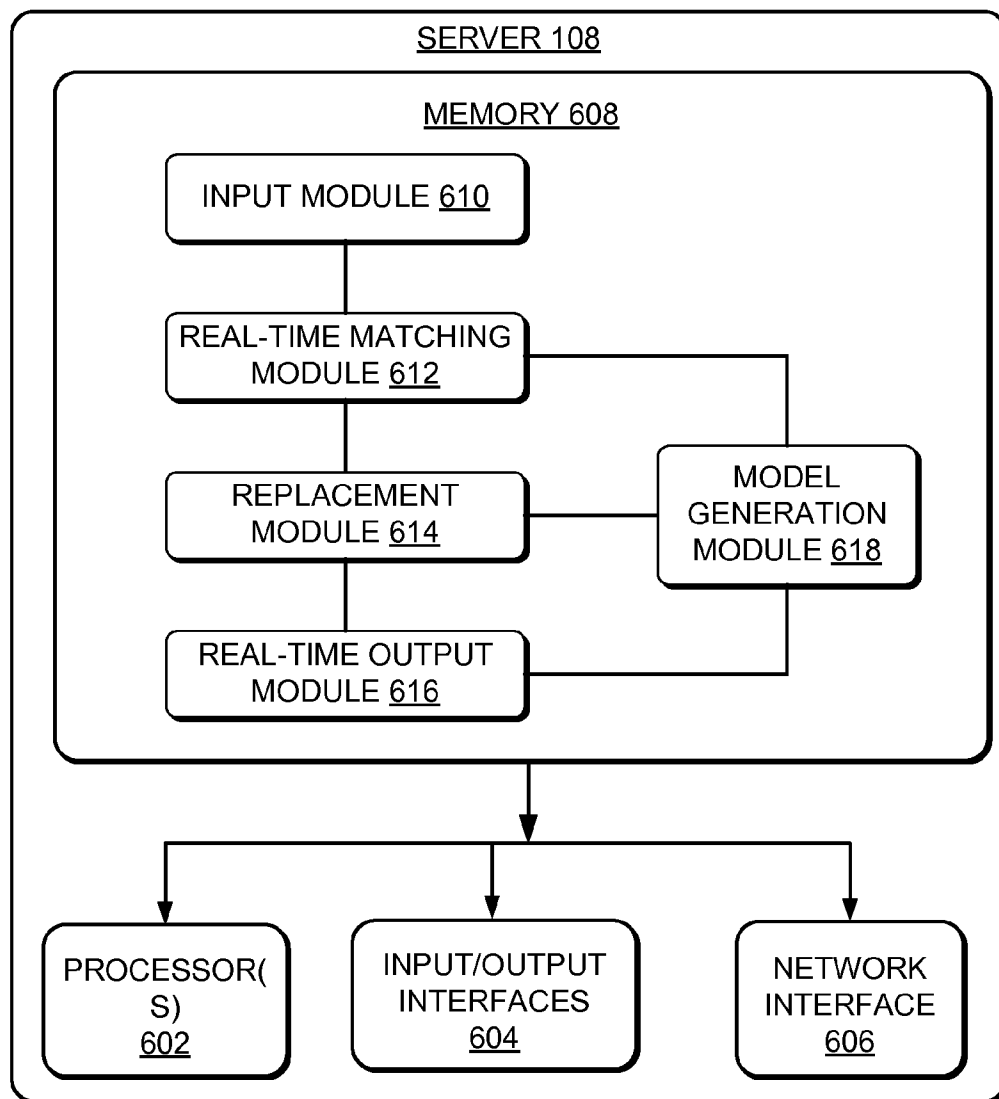


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/065096

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F17/30
ADD.

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

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

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

EP0-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2008/097955 A1 (KAPUR SHYAM [US]) 24 April 2008 (2008-04-24) paragraphs [0021] - [0038]; figures 1-3 -----	1-20
A	ALEMZADEH M ET AL: "An Efficient Method for Tagging a Query with Category Labels Using Wikipedia towards Enhancing Search Engine Results", WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY (WI-IAT), 2010 IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 31 August 2010 (2010-08-31), pages 192-195, XP031785720, ISBN: 978-1-4244-8482-9 the whole document ----- -/--	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

26 March 2013

Date of mailing of the international search report

04/04/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Moon, Timothy

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/065096

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GLOVER E J ET AL: "Improving category specific Web search by learning query modifications", PROCEEDINGS OF SAINT 2001 SYMPOSIUM ON APPLICATIONS AND THE INTERNET; 8 - 12 JANUARY 2001; SAN DIEGO, CA, USA, IEEE COMPUTER SOCIETY, LOS ALAMITOS, CA, USA, 8 January 2001 (2001-01-08), pages 23-32, XP010532795, DOI: 10.1109/SAINT.2001.905165 ISBN: 978-0-7695-0942-6 the whole document -----	1-20

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2012/065096

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2008097955	A1	24-04-2008	NONE
