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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 March 2016 (03.03.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/033237 A1

(51) International Patent Classification:
G10L 15/22 (2006.01) *G06F 17/30* (2006.01)

(21) International Application Number:
PCT/US2015/047032

(22) International Filing Date:
26 August 2015 (26.08.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/044,146 29 August 2014 (29.08.2014) US
14/812,811 29 July 2015 (29.07.2015) US

(71) Applicant: **GOOGLE INC.** [US/US]; 1600 Amphitheatre Parkway, Mountain View, California 94303 (US).

(72) Inventors: **SKOBELTSYN, Gleb**; Brandschenkestrasse 110, CH-8002 Zurich (CH). **CHEREPANOV, Evgeny A.**; Brandschenkestrasse 110, CH-8002 Zürich (CH). **BEHZADI, Behshad**; Brandschenkestrasse 110, CH-8002 Zürich (CH).

(74) Agent: **GUSTAFSON, Brian J.**; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, 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, SA, 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, ST, 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,

[Continued on next page]

(54) Title: QUERY REWRITE CORRECTIONS

(57) Abstract: Methods, systems, and apparatus, including computer programs encoded on computer storage media, for natural language processing. One of the methods includes receiving a first voice query; generating a first recognition output; receiving a second voice query; determining from a recognition of the second voice query that the second voice query triggers a correction request; using the first recognition output and the second recognition to determine a plurality of candidate corrections; scoring each candidate correction; and generating a corrected recognition output for a particular candidate correction having a score that satisfies a threshold value.

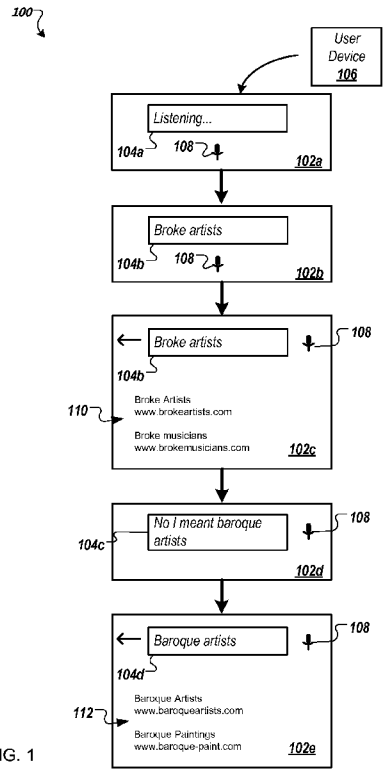


FIG. 1

WO 2016/033237 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

QUERY REWRITE CORRECTIONS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a non-provisional of and claims priority to U.S. Provisional Patent Application No. 62/044,146, filed on August 29, 2014, the entire contents of which are hereby incorporated by reference.

BACKGROUND

[0002] This specification relates to natural language processing.

[0003] The Internet provides access to a wide variety of resources, such as image files, audio files, video files, and web pages. A search system can identify resources in response to queries submitted by users and provide information about the resources in a manner that is useful to the users. The users can navigate through, e.g., select, search results to obtain information of interest.

SUMMARY

[0004] In general, one innovative aspect of the subject matter described in this specification can be embodied in methods that include the actions of receiving a first voice query; generating a first recognition output; receiving a second voice query; determining from a recognition of the second voice query that the second voice query triggers a correction request; using the first recognition output and the second recognition to determine a plurality of candidate corrections; scoring each candidate correction; and generating a corrected recognition output for a particular candidate correction having a score that satisfies a threshold value. Other embodiments of this aspect include corresponding computer systems, apparatus, and computer programs recorded on one or more computer storage devices, each configured to perform the actions of the methods. For a system of one or more computers to be configured to perform particular operations or actions means that the system has installed on it software, firmware, hardware, or a combination of them that in operation cause the system to perform the operations or actions. For one or more computer programs to be configured to perform particular operations or actions means that the one or more programs include instructions that, when executed by data processing apparatus, cause the apparatus to perform the operations or actions.

[0005] The foregoing and other embodiments can each optionally include one or more of the following features, alone or in combination. In particular, one embodiment includes all the following features in combination. Determining that the recognition of the second voice query triggers the correction request includes parsing the second voice query to detect a correction prefix. Using the first recognition output and the second recognition to determine a plurality of candidate corrected queries includes: determining a misrecognition portion of the first recognition output; and substituting the misrecognition portion with one or more candidate n-grams to form a candidate correction. The one or more candidate n-grams are obtained from one or more of: recognition of a correction portion of the second voice query; generated hypotheses of the correction portion of the second voice query; a misrecognition index of phonetically similar n-grams; or n-grams associated with a type or category associated with the correction portion of the second voice query. Each candidate corrected query is scored based at least in part on a phonetic distance between the candidate correction and the first voice input. Each candidate corrected query is scored based at least in part on a query popularity of the candidate correction. The query popularity is based on query counts in a query log that match the candidate correction. The query popularity is based at least on a change in query popularity between the first voice query and the candidate correction.

[0006] The subject matter described in this specification can be implemented in particular embodiments so as to realize one or more of the following advantages. Misrecognized voice queries can be corrected with a subsequent query without the user starting the query process over with a new query. Correcting voice queries with a corrected query improves voice search versatility. The corrected voice queries can be provided in a natural way that more closely mirrors human conversational behavior, improving user interactions with a search system.

[0007] The details of one or more embodiments of the subject matter of this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIG. 1 is an illustration of a voice search including a correction request in an example search user interface.

[0009] FIG. 2 is a block diagram of a system for providing corrections to voice queries.

[0010] FIG. 3 is a flow diagram of an example process for correcting an input voice query.

[0011] FIG. 4 is a block diagram illustrating example techniques for generating candidate correction.

[0012] FIG. 5 is a block diagram illustrating example factors in generating a score for a candidate correction.

[0013] Like reference numbers and designations in the various drawings indicate like elements.

DETAILED DESCRIPTION

[0014] FIG. 1 is an illustration 100 of a voice search including a correction request in an example search user interface. A search system provides a search user interface 102a for displaying on a user device 106. The user interface 102a includes a search box 104a for receiving a query and a microphone icon 108 for receiving a first voice query from a user. The user inputs a first voice query, e.g., by selecting the microphone icon 108 in the user interface 102a.

[0015] In a user session, the search system receives the first voice query from a user. In particular, the user provides a voice input of "Baroque artists." In response to the first voice query, the user interface 102b is presented including a display of a first recognition output generated from the first voice query in a search box 104b. The first recognition output is [broke artists]. For example, a speech-to-text process can be performed on the first voice query to generate a purported transcription as the corresponding recognition output. The search system can provide the transcription to the user device 106 for display in the user interface 102b.

[0016] Furthermore, the recognition output can be used by a search system. The search system provides search results 110 responsive to the query [broke artists] for display in user interface 102c. For example, the search results 110 can include links to particular resources determined by the search system to be responsive to the recognized query. However, the recognized query was misrecognized. The user therefore inputs a second voice query, e.g., by selecting the microphone icon 108 in the user interface 102c, e.g., "no I meant baroque artists."

[0017] In response to the second voice query, a user interface 102d is presented including a display of a second recognition output from the second voice input in a search box 104c. The second recognized query is [No I meant baroque artists]. The search system determines that the second voice query is a correction of the first voice query, for example, based on recognizing "no I meant" as a correction prefix.

[0018] The search system determines a corrected recognition output, [baroque artists], which is presented in a search box 104d of a user interface 102e. Furthermore, in response to the corrected recognition output, the search system provides search results 112 responsive to corrected query [baroque artists] for presentation in the user interface 102e. For example, the search results 112 can include links to particular resources determined by the search system to be responsive to the corrected query.

[0019] FIG. 2 is a block diagram of a system 200 for providing corrections to voice queries. The system 200 provides search results relevant to submitted queries as can be implemented in an internet, an intranet, or another client and server environment. The system 200 is an example of an information retrieval system in which the systems, components, and techniques described below can be implemented.

[0020] A user 202 can interact with a search system 214 through a client device 204. For example, the client 204 can be a computer coupled to the search system 214 through a local area network (LAN) or wide area network (WAN), e.g., the Internet. In some implementations, the search system 214 and the client device 204 can be one machine. For example, a user can install a desktop search application on the client device 204. The client device 204 will generally include a random access memory (RAM) 206 and a processor 208.

[0021] The user 202 can submit voice queries 210 to a search engine 230 within a search system 214. When the user 202 submits a voice query 210, the voice query 210 is transmitted through a network to the search system 214. The search system 214 can be implemented as, for example, computer programs running on one or more computers in one or more locations that are coupled to each other through a network.

[0022] The search system 214 includes a voice recognition engine 254. The voice recognition engine receives the voice query 210 and transcribes the voice query to a recognized query, e.g., using suitable text-to-speech techniques. In some implementations, the recognized query is returned to the client device 204 for presentation to the user 202. Additionally, the recognized query is provided to the search engine 230.

[0023] The search system 214 further includes an index database 222 and a search engine 230. The search system 214 responds to an input query by generating search results 228, which are transmitted through the network to the client device 204 in a form that can be presented to the user 202 (e.g., as a search results web page to be displayed in a web browser running on the client device 204).

[0024] When the recognized query determined from the voice query 210 is received by the search engine 230, the search engine 230 identifies resources that match, or are responsive to, the query 210. The search engine 230 will generally include an indexing engine 220 that indexes resources (e.g., web pages, images, or news articles on the Internet) found in a corpus (e.g., a collection or repository of content), an index database 222 that stores the index information, and a ranking engine 252 (or other software) to rank the resources that match the query 210. The indexing and ranking of the resources can be performed using conventional techniques. The search engine 130 can transmit the search results 228 through the network to the client device 204 for presentation to the user 202.

[0025] The search system also includes a correction engine 256. The correction engine 256 corrects the recognized query in response to a correction input from the user 202. In particular, the correction engine 256 can determine that a received voice query 210 includes a correction request, determine candidate corrected queries, score the candidate correction queries, and if a threshold is satisfied, select a particular correction as a corrected recognition output. The selected corrected output is then provided to the search engine 230 as a corrected query and used identify resources that are responsive to the corrected query.

[0026] FIG. 3 is a flowchart of an example process 300 for correcting an input voice query. For convenience, the process 300 will be described as being performed by a system of one or more computers, located in one or more locations, and programmed appropriately in accordance with this specification. For example, a search system, e.g., the search system 214 of FIG. 2, appropriately programmed, can perform the process 300.

[0027] The system receives a first voice query from a user device 302. For example, a search user interface of the user device can include a microphone indicator. When selected by the user, a microphone of the user device captures voice input from the user. The voice input is then transmitted by the user device to the system. The first voice query 302 can be, for example, “who is the president of France?”

[0028] The system generates a recognition output 304. The recognition output is a transcription of the received first voice query. For example, a suitable speech to text technique can be used to provide voice recognition of the first voice query and convert it into a recognized text output. In some implementations, the speech to text technique can correct or compensate for errors in the first voice query, e.g., based on spelling and/or grammar rules. The recognition output is provided to the user device, for example, for display in the search interface. The recognition output can be displayed, for example, to indicate the system's recognition of the first voice query. The user can then examine the presented recognition output to determine whether the system correctly recognized the first voice query. For example, the first voice query "who is the president of France" may be recognized as [who is the president of friends]. Here the word "France" was misrecognized as "friends."

[0029] Additionally, the system can perform a search using the recognition output as a search query. Search results responsive to the search query can be provided to the user for display in the search interface.

[0030] The system receives a second voice query from the user device 306. The second voice query is received in a similar manner from the user device as the first voice query. In particular, the second voice query can be a correction query. For example, the second voice query in response to the recognition output "who is the president of friends" may be "no I meant France."

[0031] The system determines whether a correction request is triggered by the second voice query 308. The system determines whether a correction request is triggered by parsing the query according to different correction request grammars. For example, the recognition of the second voice query can be parsed to determine whether the query starts from a "correction prefix" e.g., "no I mean…" indicating that the query is a correction of a prior query rather than a new query. In some implementations, the system uses voice recognition hypotheses and fuzzy matching in phonetic space to detect correction prefixes indicating a correction request.

[0032] Example grammars that indicate a correction request include:

[0033] 1. Repeating a part of the query containing the misrecognized word(s):

[0034] Query 1: [who is president of friends]

[0035] Query 2: [president of France]

[0036] 2. Using a *correction command* followed by a part of the query containing the misrecognized word(s):

[0037] Query 1: [who is president of friends]

[0038] Query 2: [*No*, France] or

[0039] Query 2: [*I said* president of France] or

[0040] Query 2: [*No I mean* France]

[0041] 3. Using a *correction replace command* with both the misrecognized and correct word(s):

[0042] Query 1: [who is president of friends]

[0043] Query 2: [*No, not friends*, France] or

[0044] Query 2: [I said France *not friends*]

[0045] 4. Using a *correction command* followed by a clarification or disambiguation phrase for the misrecognized word(s):

[0046] By type or category:

[0047] Query 2: [*I meant* country France] or

[0048] Query 2: [*No I mean* lynx the animal]

[0049] By a phrase that contains the misrecognized word(s) and a disambiguation phrase:

[0050] Query 2: [*I meant* be like in "honey bee"]

[0051] Query 2: [*No I mean* lynx as in "chain links"]

[0052] 5. Using a semantic correction, for example, a *correction command* followed by a type, category, or source for the misrecognized word(s):

[0053] Query 1: [who is president of friends]

[0054] Query 2: *[I meant* the country]

[0055] Query 1: [best fries in Russells]

[0056] Query 2: [*No I mean* the city in Europe]

[0057] 6. Using [no it is spelled] command followed by the spelling of the misrecognized word:

[0058] Query 1: [Barack Obama]

[0059] Query 2: [*No it's spelled* O-B-A-M-A] or

[0060] Query 2: [*It's spelled* Oslo Boston Atlanta Miami Atlanta]

[0061] 7. Using a *correction command* followed by a "spelling clarification" for the misrecognized word:

[0062] Query 2: [I meant lynx with x] or

[0063] Query 2: [No I mean Berklee with double-E]

[0064] 8. Using a *correction command* followed by a part of the query containing the misrecognized word(s) which is/are misrecognized again:

[0065] Query 1: [who is president of friends]

[0066] Query 2: [*No* Franz]

[0067] In response to the system determining that a correction request is not triggered, the system sends the recognition output to the client device and provides one or more search results responsive to a query corresponding to the recognition output (316).

[0068] In response to the system determining that a correction request is triggered, the system determines candidate corrections to the recognition output of the first voice query (310). To form each candidate correction, an n-gram is substituted into the misrecognition of the recognition output. For example, the misrecognition [who is president of friends] can include as a candidate correction the substitution of "France" for "friends" to generate candidate correction [who is president of France]. One or more different techniques can be used to generate the candidate corrections, as described with respect to FIG. 4.

[0069] FIG. 4 is a block diagram 400 illustrating example techniques for generating candidate corrected queries. The candidate generation techniques use one or more of the recognition of the first voice query 402 and the recognition of the second voice query 404 as input and generates candidate corrections 406 as outputs.

[0070] Candidate corrections can be generated using a recognition from the correction request 408, i.e., the second voice query 404. In particular, a candidate correction can be obtained by substituting the misrecognition with the correction from the second voice input. For example, if the recognition output of the first voice query is "who is president of friends" and the second voice input is recognized as [no I meant France] the portion following the correction prefix [no I meant] is used as the candidate substitution n-gram into the first recognition output. In this example, the candidate substitution n-gram "France" can be substituted into the original recognition. This results in candidate corrections [who is president of France], [who is president France], [who is the France], and [who is France].

[0071] Candidate corrections can be generated using various hypotheses of the correction request 410. The correction request may also be misrecognized. For example, [who is president of friends] can be followed by the correction request recognized as [no, I said Franz,] where "France" was still misrecognized, this time as "Franz." Additional

recognition hypotheses for the second voice input can then be used to generate additional candidate corrections.

[0072] In the above example, additional hypotheses for recognizing the second voice input of "no I said France" can include [no I said Franz], [no I said France], [no I sat France], [noah sad friends] and so on. These hypotheses, generated e.g., from speech-to-text recognition, are used, as described with the recognition from the correction request 408, to generate candidate corrections by substituting n-grams of the various hypotheses into the misrecognized output for the first voice query including, for example, [who is the president of Franz], [who is the president of France], etc.

[0073] Candidate corrections can be generated using an obtained misrecognition index 412. The misrecognition index can be obtained by the system for use in identifying candidate corrections. The misrecognition index provides a lookup of possible phonetically similar values to a part in question with a give context or type. In some implementations, the misrecognition index is generated from typed queries and n-grams and maintains pre-computed top-N lists for keys derived from phonetic representation. The top-N lists are re-ranked based on phonetic similarity of actual phonetic representations. The phonetically similar values can be substituted into the misrecognition to generate candidate corrections. The n-grams can be ranked, for example, based on how often a particular first n-gram is misrecognized as a particular replacement second n-gram in voice query logs and how popular the query having the first n-gram replaced by the second n-gram is.

[0074] For example, the misrecognition "friends" with the left context "president of" can be identified in the index as being phonetically similar to, for example, "France," "Iran", and "finance." The misrecognition index can also provide a lookup of phonetically similar values to a part in question based on a particular type. For example, the misrecognition "mouse" can be phonetically similar to terms of the type "bug" including "moth," "mantis," "mantid," etc. The misrecognition index can be generated using obtained semantic relationships between entities, e.g., to determine that "France" and "Iran" have a same entity type of "Country."

[0075] Candidate corrections can be generated using spelling analysis 414. In particular, the second voice query can recite the spelling of the misrecognition. For example, the correction request can be recognized as [No it's spelled B-E-L-T-R-A-N]. The system can assemble the spelling in to a single word and use it as a substitution into the misrecognition.

[0076] Candidate corrections can be generated using type or category analysis 416. For example, the first voice query can be "who is the president of France" recognized as [who is the president of friends]. The correction request can be recognized as [no I meant the country]. In this case, instead of repeating the correct term, France, the user indicated a category to which the correct term belongs. Using the correction request as the substitution (408), one candidate correction would be [who is the president of country]. However, the system can also recognize that the type/category that the user wants to correct is "country." For example, and index or other collection of terms known to represent types or categories can be used to identify instances where the correction input is a reference to a type or category of information. The type/category index can be used to identify entities with type "country" that are phonetically similar to "friends." These similar entities of type "country" can then be substituted to generate additional candidate corrections.

[0077] Referring back to FIG. 3, the candidate corrections are scored (312). Each candidate correction is assigned a score value based on one or more applied scoring techniques. Each candidate correction includes information on the replacement source and target as well as an interpretation of the correction request based, e.g., on the grammar used. One or more different techniques can be used to score candidate corrections, as described with respect to FIG. 5.

[0078] FIG. 5 is a block diagram 500 of example factors in generating a score for a candidate corrected query. The score 504 for a candidate corrected query 502 can be generated using one or more scoring factors. For example, multiple scoring factors can be used in a scoring formula. The scoring formula can optionally be weighted, giving more scoring weight to particular scoring factors.

[0079] A candidate correction can be scored based on a confidence in the grammar matching 506. This can include a determination of a probability that the second voice query represents a correction request based on the parsed specific grammar. The confidence in the grammar matching can also include grammar specific scoring, for example, how well a particular "type" matches a term in the correction request. In some implementations, a higher weight is given to grammars that are more general as compared to specific grammars. In some implementations, the grammar based weighting is used to tune how aggressively the system directly substitutes the correction vs. using the grammar type, e.g., using "country" as a type rather than substituting "country" for the misrecognized query following the correction [no I meant the country].

[0080] The candidate correction can be scored based on a measure of replacement phonetic quality 508. The replacement phonetic quality can be based on the phonetic similarity between n-grams representing the candidate correction n-gram and the n-gram in the misrecognition being substituted for. In particular, this is a measure of the probability that a user provided a voice input that was recognized as the misrecognition. A phonetic distance can be calculated between the two n-grams to provide a measure of the phonetic similarity. Various suitable measures for similarity can be used including, for example, a phonetic similarity function using phonetic edit distance.

[0081] The candidate correction can be scored based on a measure of query quality 510. The measure of query quality represents a probability that the user wanted to issue the candidate correction as a query. The probability can be based on one or more of a query popularity, a pattern of the query, or whether the query triggers an answer box or other direct search response. The query popularity can be measured based at least in part on counts of how many times a matching or similar query has been issued, e.g., using logs. The pattern of the query can be used to determine a semantic quality of the query. In some other implementations, the query quality can also include length of the candidate correction as a factor. In particular, longer candidate corrections having a threshold popularity value can be scored higher than shorter candidate corrections having a similar popularity.

[0082] The candidate correction can be scored based on a measure of replacement query quality 512. Replacement query quality represents how the query would improve if the candidate correction replaces the original misrecognized query. Replacement query quality can be based on, for example, popularity of the original query vs. the rewritten query. The popularity can be estimated statistically for queries that are not found in query logs. The replacement query quality can also be based on search results generated for each respective query. A quality measure for results can be compared between the queries.

[0083] The scoring factors can be combined to generate the overall candidate score 504 for each candidate correction 502.

[0084] Referring back to FIG. 3, the system provides a highest scoring candidate correction as a corrected recognition output if a threshold score value is satisfied 314. In particular, a threshold score value can be established such that an uncorrected recognition output is provided in response to score values that do not satisfy the threshold. For

example, the scores can be such that there is not a high degree of confidence in any of the candidate corrections.

[0085] The highest scoring candidate correction becomes the corrected recognition output and can be provided for display on the client device. Additionally, the corrected recognition output can be used as a search query. One or more search results responsive to the search query can be provided to the user device for display, e.g., in a search interface. Additionally, in some implementations, the search query triggers a particular answer box. The answer box provides a specific entity or other information in response to the query. For example, for the corrected recognition output of [who is the president of France], the system can provide an answer box that indicates the president of France is François Hollande. The answer box can be supplemented by one or more search results.

[0086] Similarly, in this specification the term "engine" will be used broadly to refer to a software based system or subsystem that can perform one or more specific functions. Generally, an engine will be implemented as one or more software modules or components, installed on one or more computers in one or more locations. In some cases, one or more computers will be dedicated to a particular engine; in other cases, multiple engines can be installed and running on the same computer or computers.

[0087] Embodiments of the subject matter and the functional operations described in this specification can be implemented in digital electronic circuitry, in tangibly-embodied computer software or firmware, in computer hardware, including the structures disclosed in this specification and their structural equivalents, or in combinations of one or more of them. Embodiments of the subject matter described in this specification can be implemented as one or more computer programs, i.e., one or more modules of computer program instructions encoded on a tangible non-transitory storage medium for execution by, or to control the operation of, data processing apparatus. The computer storage medium can be a machine-readable storage device, a machine-readable storage substrate, a random or serial access memory device, or a combination of one or more of them. Alternatively or in addition, the program instructions can be encoded on an artificially-generated propagated signal, e.g., a machine-generated electrical, optical, or electromagnetic signal, that is generated to encode information for transmission to suitable receiver apparatus for execution by a data processing apparatus.

[0088] The term "data processing apparatus" refers to data processing hardware and encompasses all kinds of apparatus, devices, and machines for processing data,

including by way of example a programmable processor, a computer, or multiple processors or computers. The apparatus can also be, or further include, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application-specific integrated circuit). The apparatus can optionally include, in addition to hardware, code that creates an execution environment for computer programs, e.g., code that constitutes processor firmware, a protocol stack, a database management system, an operating system, or a combination of one or more of them.

[0089] A computer program, which may also be referred to or described as a program, software, a software application, a module, a software module, a script, or code, can be written in any form of programming language, including compiled or interpreted languages, or declarative or procedural languages; and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A program may, but need not, correspond to a file in a file system. A program can be stored in a portion of a file that holds other programs or data, e.g., one or more scripts stored in a markup language document, in a single file dedicated to the program in question, or in multiple coordinated files, e.g., files that store one or more modules, sub-programs, or portions of code. A computer program can be deployed to be executed on one computer or on multiple computers that are located at one site or distributed across multiple sites and interconnected by a data communication network.

[0090] The processes and logic flows described in this specification can be performed by one or more programmable computers executing one or more computer programs to perform functions by operating on input data and generating output. The processes and logic flows can also be performed by special purpose logic circuitry, e.g., an FPGA or an ASIC, or by a combination of special purpose logic circuitry and one or more programmed computers.

[0091] Computers suitable for the execution of a computer program can be based on general or special purpose microprocessors or both, or any other kind of central processing unit. Generally, a central processing unit will receive instructions and data from a read-only memory or a random access memory or both. The essential elements of a computer are a central processing unit for performing or executing instructions and one or more memory devices for storing instructions and data. The central processing unit and the memory can be supplemented by, or incorporated in, special purpose logic circuitry. Generally, a computer will also include, or be operatively coupled to receive data from or

transfer data to, or both, one or more mass storage devices for storing data, e.g., magnetic, magneto-optical disks, or optical disks. However, a computer need not have such devices. Moreover, a computer can be embedded in another device, e.g., a mobile telephone, a personal digital assistant (PDA), a mobile audio or video player, a game console, a Global Positioning System (GPS) receiver, or a portable storage device, e.g., a universal serial bus (USB) flash drive, to name just a few.

[0092] Computer-readable media suitable for storing computer program instructions and data include all forms of non-volatile memory, media and memory devices, including by way of example semiconductor memory devices, e.g., EPROM, EEPROM, and flash memory devices; magnetic disks, e.g., internal hard disks or removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks.

[0093] Control of the various systems described in this specification, or portions of them, can be implemented in a computer program product that includes instructions that are stored on one or more non-transitory machine-readable storage media, and that are executable on one or more processing devices. The systems described in this specification, or portions of them, can each be implemented as an apparatus, method, or electronic system that may include one or more processing devices and memory to store executable instructions to perform the operations described in this specification.

[0094] To provide for interaction with a user, embodiments of the subject matter described in this specification can be implemented on a computer having a display device, e.g., a CRT (cathode ray tube) or LCD (liquid crystal display) monitor, for displaying information to the user and a keyboard and a pointing device, e.g., a mouse or a trackball, by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback provided to the user can be any form of sensory feedback, e.g., visual feedback, auditory feedback, or tactile feedback; and input from the user can be received in any form, including acoustic, speech, or tactile input. In addition, a computer can interact with a user by sending documents to and receiving documents from a device that is used by the user; for example, by sending web pages to a web browser on a user's device in response to requests received from the web browser.

[0095] Embodiments of the subject matter described in this specification can be implemented in a computing system that includes a back-end component, e.g., as a data server, or that includes a middleware component, e.g., an application server, or that includes a front-end component, e.g., a client computer having a graphical user interface

or a web browser through which a user can interact with an implementation of the subject matter described in this specification, or any combination of one or more such back-end, middleware, or front-end components. The components of the system can be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network (LAN) and a wide area network (WAN), e.g., the Internet.

[0096] The computing system can include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other. In some embodiments, a server transmits data, e.g., an HTML page, to a user device, e.g., for purposes of displaying data to and receiving user input from a user interacting with the user device, which acts as a client. Data generated at the user device, e.g., a result of the user interaction, can be received from the user device at the server.

[0097] While this specification contains many specific implementation details, these should not be construed as limitations on the scope of any invention or on the scope of what may be claimed, but rather as descriptions of features that may be specific to particular embodiments of particular inventions. Certain features that are described in this specification in the context of separate embodiments can also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment can also be implemented in multiple embodiments separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claimed combination may be directed to a subcombination or variation of a subcombination.

[0098] Similarly, while operations are depicted in the drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system modules and components in the embodiments described above should not be understood as requiring such separation in all embodiments, and it should be understood that the

described program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

[0099] Particular embodiments of the subject matter have been described. Other embodiments are within the scope of the following claims. For example, the actions recited in the claims can be performed in a different order and still achieve desirable results. As one example, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In some cases, multitasking and parallel processing may be advantageous.

[00100] What is claimed is:

CLAIMS

1. A method comprising:

receiving a first voice query;

generating a first recognition output;

receiving a second voice query;

determining from a recognition of the second voice query that the second voice query triggers a correction request;

using the first recognition output and the second recognition to determine a plurality of candidate corrections;

scoring each candidate correction; and

generating a corrected recognition output for a particular candidate correction having a score that satisfies a threshold value.

2. The method of claim 1, wherein determining that the recognition of the second voice query triggers the correction request includes parsing the second voice query to detect a correction prefix.

3. The method of claim 1, wherein using the first recognition output and the second recognition to determine a plurality of candidate corrected queries comprises:

determining a misrecognition portion of the first recognition output; and

substituting the misrecognition portion with one or more candidate n-grams to form a candidate correction.

4. The method of claim 3, wherein the one or more candidate n-grams are obtained from one or more of:

recognition of a correction portion of the second voice query;

generated hypotheses of the correction portion of the second voice query;

a misrecognition index of phonetically similar n-grams; or

n-grams associated with a type or category associated with the correction portion of the second voice query.

5. The method of claim 1, wherein each candidate corrected query is scored based at least in part on a phonetic distance between the candidate correction and the first voice input.

6. The method of claim 1, wherein each candidate corrected query is scored based at least in part on a query popularity of the candidate correction.

7. The method of claim 6, wherein the query popularity is based on query counts in a query log that match the candidate correction.

8. The method of claim 6, wherein the query popularity is based at least on a change in query popularity between the first voice query and the candidate correction.

9. A system comprising:

a user device; and

one or more computers configured to interact with the user device and to perform operations comprising:

receiving a first voice query from the user device;

generating a first recognition output;

receiving a second voice query from the user device;

determining from a recognition of the second voice query that the second voice query triggers a correction request;

using the first recognition output and the second recognition to determine a plurality of candidate corrections;

scoring each candidate correction; and

generating a corrected recognition output for a particular candidate correction having a score that satisfies a threshold value.

10. The system of claim 9, wherein determining that the recognition of the second voice query triggers the correction request includes parsing the second voice query to detect a correction prefix.

11. The system of claim 9, wherein using the first recognition output and the second recognition to determine a plurality of candidate corrected queries comprises:

determining a misrecognition portion of the first recognition output; and

substituting the misrecognition portion with one or more candidate n-grams to form a candidate correction.

12. The system of claim 11, wherein the one or more candidate n-grams are obtained from one or more of:

recognition of a correction portion of the second voice query;

generated hypotheses of the correction portion of the second voice query;

a misrecognition index of phonetically similar n-grams; or

n-grams associated with a type or category associated with the correction portion of the second voice query.

13. The system of claim 9, wherein each candidate corrected query is scored based at least in part on a phonetic distance between the candidate correction and the first voice input.

14. The system of claim 9, wherein each candidate corrected query is scored based at least in part on a query popularity of the candidate correction.

15. The system of claim 14, wherein the query popularity is based on query counts in a query log that match the candidate correction.

16. The system of claim 14, wherein the query popularity is based at least on a change in query popularity between the first voice query and the candidate correction.

17. One or more computer-readable storage media encoded with instructions that, when executed by one or more computers, cause the one or more computers to perform operations comprising:

receiving a first voice query;

generating a first recognition output;

receiving a second voice query;

determining from a recognition of the second voice query that the second voice query triggers a correction request;

using the first recognition output and the second recognition to determine a plurality of candidate corrections;

scoring each candidate correction; and

generating a corrected recognition output for a particular candidate correction having a score that satisfies a threshold value.

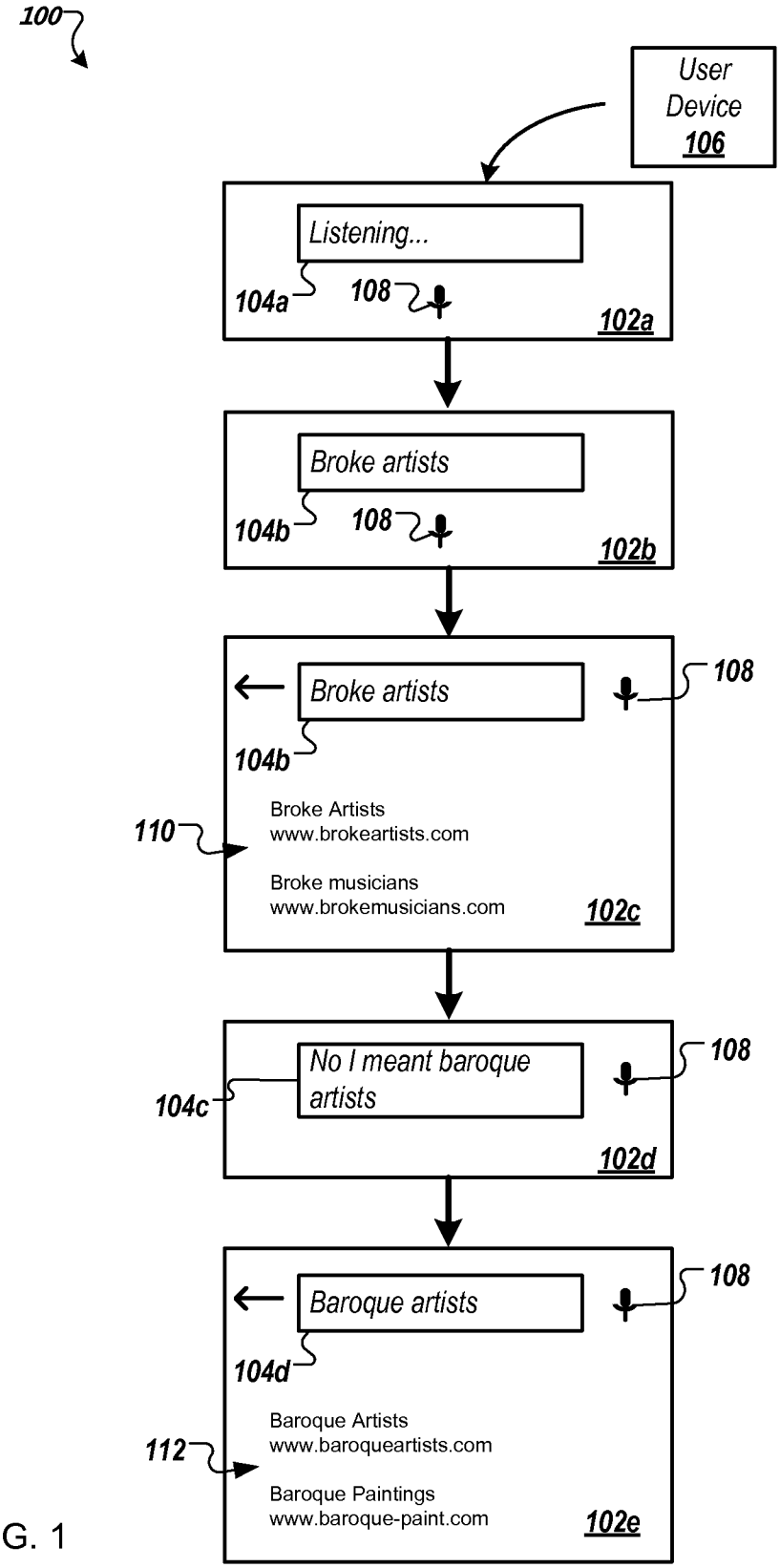

100
User Device 106
Listening...
104a
108
102a
Broke artists
104b
108
102b
Broke artists
104b
108
110
Broke Artists
www.brokeartists.com
Broke musicians
www.brokemusicians.com
102c
No I meant baroque artists
104c
108
102d
Baroque artists
104d
108
112
Baroque Artists
www.baroqueartists.com
Baroque Paintings
www.baroque-paint.com
102e

FIG. 1

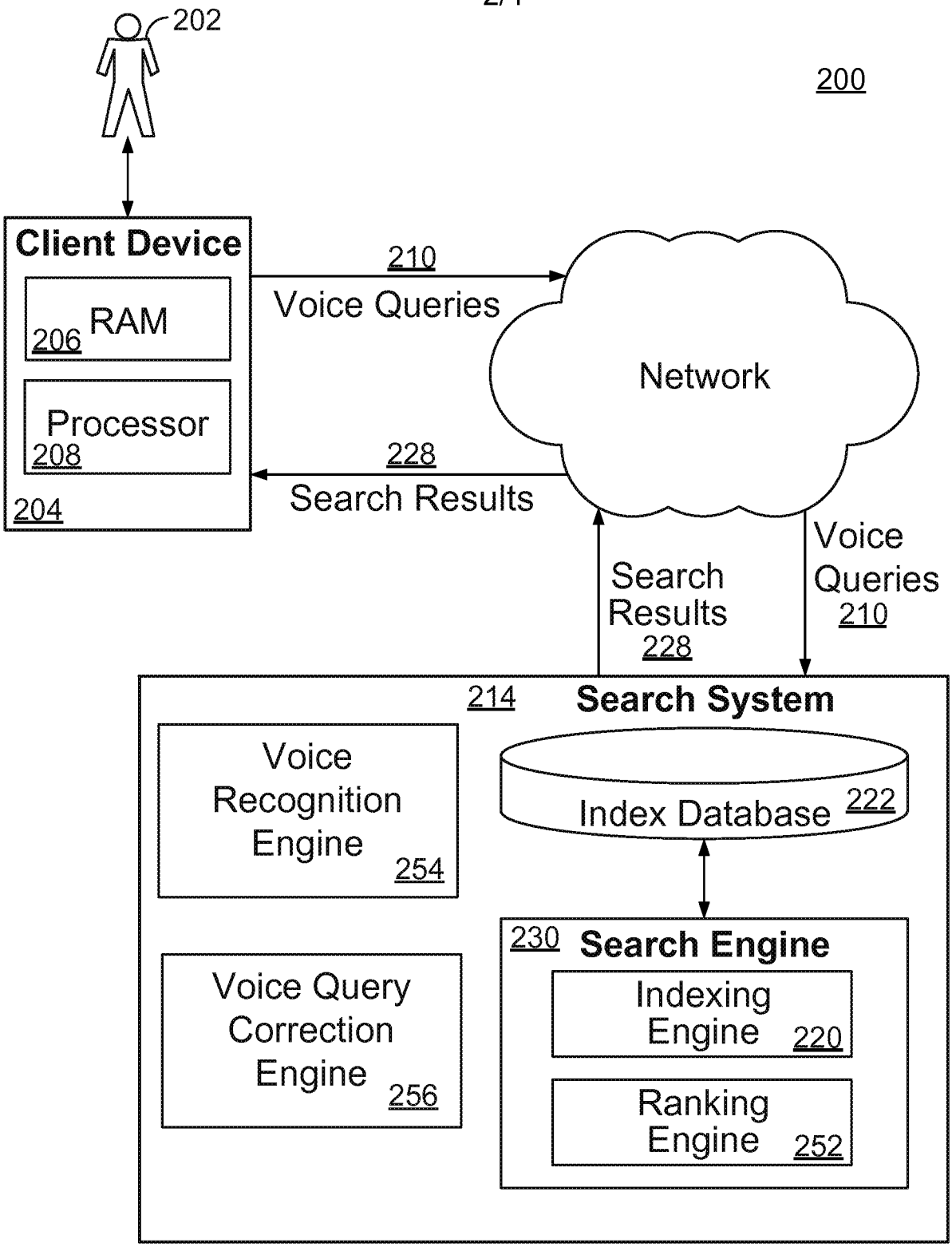

202
200
Client Device
206 RAM
Processor
208
204
210
Voice Queries
228
Search Results
Network
Search
Results
228
Voice
Queries
210
214
Search System
Voice
Recognition
Engine
254
Index Database
222
230
Search Engine
Indexing
Engine
220
Ranking
Engine
252
Voice Query
Correction
Engine
256

FIG. 2

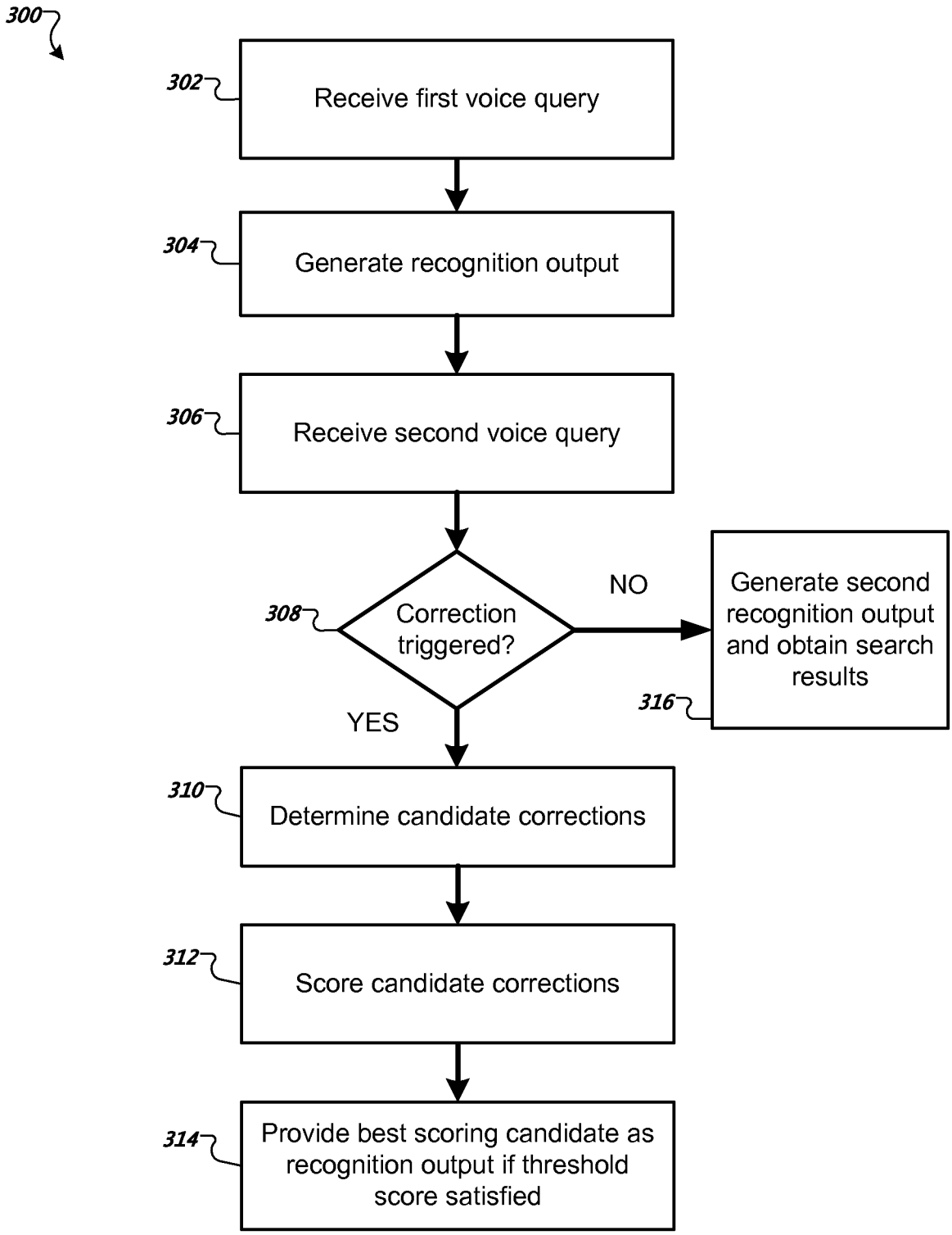

300
302
Receive first voice query
304
Generate recognition output
306
Receive second voice query
308
Correction triggered?
NO
Generate second recognition output and obtain search results
316
YES
310
Determine candidate corrections
312
Score candidate corrections
314
Provide best scoring candidate as recognition output if threshold score satisfied

FIG. 3

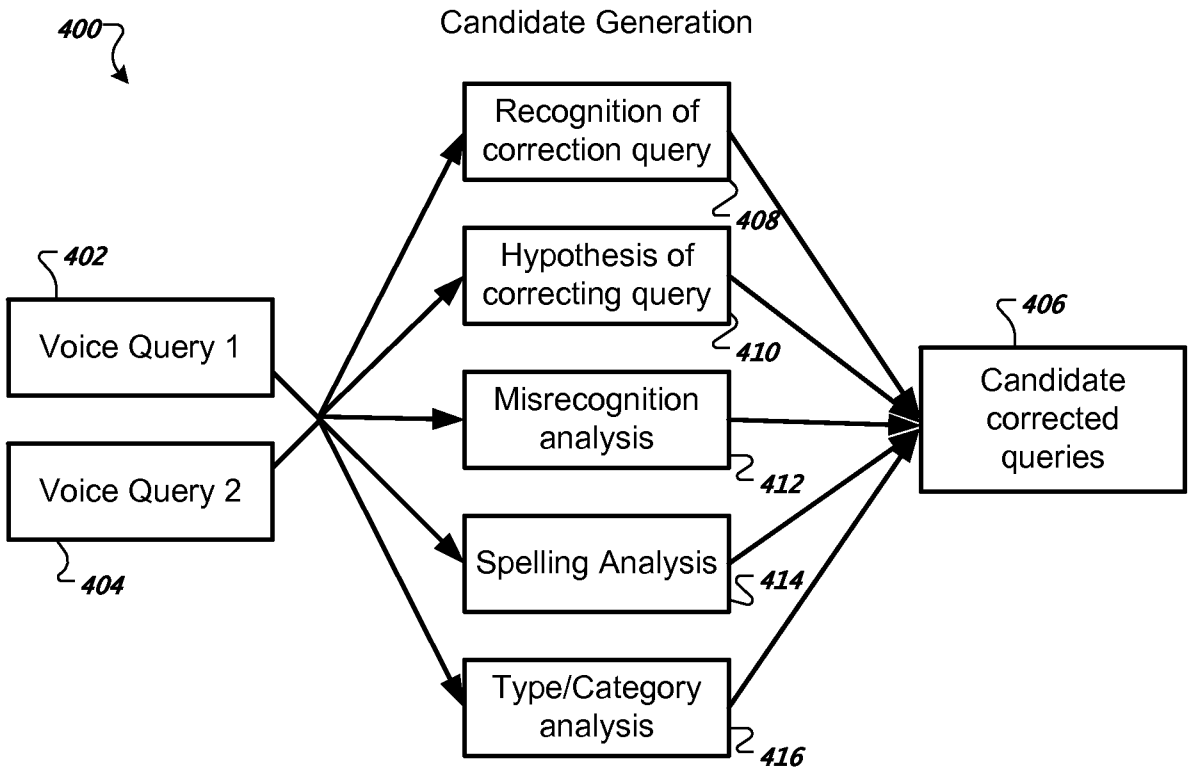

400
Candidate Generation
Recognition of correction query
408
Hypothesis of correcting query
410
Misrecognition analysis
412
Spelling Analysis
414
Type/Category analysis
416
402
Voice Query 1
Voice Query 2
404
406
Candidate corrected queries

FIG. 4

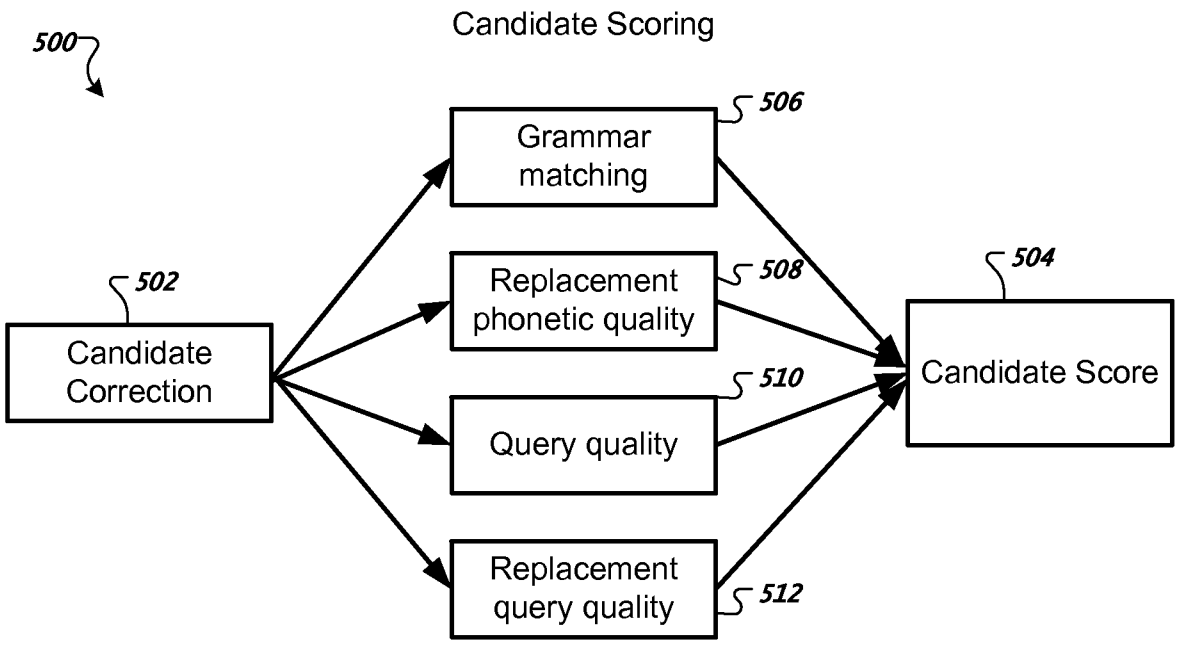

500
Candidate Scoring
Grammar matching
506
Replacement phonetic quality
508
Query quality
510
Replacement query quality
512
502
Candidate Correction
504
Candidate Score

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/047032

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/22 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)
G10L 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)
EPO-Internal, COMPENDEX, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 189 203 A2 (L & H HOLDINGS USA INC [US]) 20 March 2002 (2002-03-20)	1-17
Y	D1,: In particular Figure 17 and paragraphs 0017, 0087, 0097, 0099, 0110, 0114, 0125, 0127, 0128, 0130, 0131, 0145, 0156.	1-17

X	US 6 064 959 A (YOUNG JONATHAN HOOD [US] ET AL) 16 May 2000 (2000-05-16) In particular, column 16, line 26 to column 25, line 25	1-17

X	US 5 794 189 A (DRAGON SYSTEMS, INC) 11 August 1998 (1998-08-11) In particular, column 9, line 45 to column 12, line 53	1-17

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] 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	Date of mailing of the international search report
28 October 2015	09/11/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Thean, Andrew

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/047032

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2007/061317 A1 (RAMER JOREY [US] ET AL) 15 March 2007 (2007-03-15) In particular paragraphs 0294, 0307, 0308, 0519 and Figures 1 and 6 -----	1-17

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2015/047032

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1189203 A2	20-03-2002	AT 326754 T	15-06-2006
		DE 60119643 T2	01-02-2007
		EP 1189203 A2	20-03-2002
US 6064959 A	16-05-2000	EP 0867858 A2	30-09-1998
		EP 1321926 A1	25-06-2003
		US 6064959 A	16-05-2000
US 5794189 A	11-08-1998	DE 69632517 D1	24-06-2004
		DE 69632517 T2	12-05-2005
		EP 0773532 A2	14-05-1997
		EP 1341156 A1	03-09-2003
		US 5794189 A	11-08-1998
US 2007061317 A1	15-03-2007	NONE	