



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2015/014300

(22) International Filing Date:
3 February 2015 (03.02.2015)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **HEWLETT-PACKARD DEVELOPMENT COMPANY, LP** [US/US]; 11445 Compaq Center Drive W, Houston, Texas 77070 (US).

(72) Inventor: **FORMAN, George**; 10993 Fraser Rd. SE, Port Orchard, Washington 98367-9703 (US).

(74) Agents: **ESTEBAN, Michelle** et al.; Hewlett-Packard Company, Intellectual Property Administration, 3404 E Harmony, Rd, Mail Stop 35, Fort Collins, Colorado 80528-9599 (US).

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.

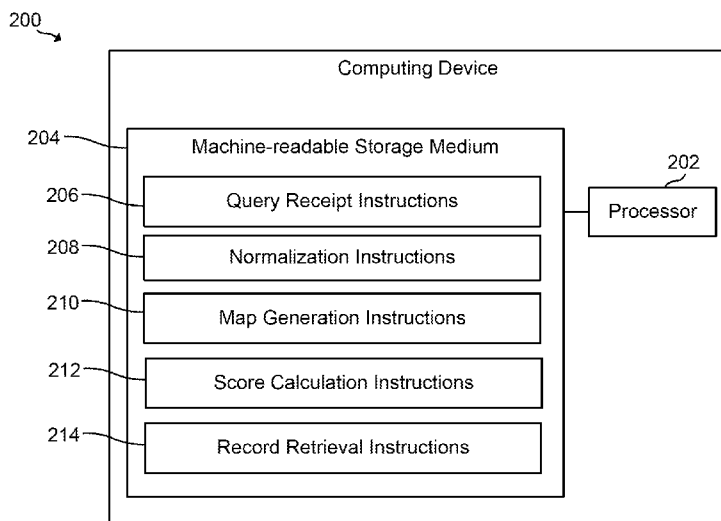
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, 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))

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

(54) Title: RECORDS BASED ON BIT-SHIFTING



(57) Abstract: Example implementations relate to records based on bit-shifting. For example, a computing device may include at least one processor to receive a query, create a normalized query, and generate a mapping of the normalized query for each record in a set of records in a database using a state array, the mapping being based on bit-shifting of bits associated with similarities between each character in the normalized query and each character in a field of the respective record in the set of records, each successive level of the state array indicating a stronger match between the query and the respective record than an immediately prior level in the state array. The processor may calculate a score associated with each record based on the respective mapping and may provide a subset of the set of records in response to the query based on the score associated with each record.

FIG. 2

RECORDS BASED ON BIT-SHIFTING

BACKGROUND

[0001] Many entities (e.g., enterprises, organizations, computer applications, etc.) utilize databases for storage of data relating to the entities. For example, a business may maintain a database of customer information, and the customer information may be accessed by querying the database. Data stored in these databases may be accessed and updated for various purposes.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] Some examples of the present application are described with respect to the following figures:

[0003] FIG. 1 is an interface diagram of an example user interface for accessing database records based on bit-shifting;

[0004] FIG. 2 is a block diagram of a computing device for providing database records based on bit-shifting; and

[0005] FIG. 3 is a flowchart illustrating an example method of returning database records based on bit-shifting.

DETAILED DESCRIPTION

[0006] As described above, data stored in a database may be accessed and updated for various purposes. For example, a company may receive phone calls from customers for various reasons (e.g., new orders from customers, service-related calls from customers, etc.), and information associated with the customer may be entered in the company's database. However, in some examples, the customer may already have an existing database record in the company's database, and entering the customer's information as a new record may be redundant.

[0007] The technology disclosed herein allows an employee of a company to efficiently access existing records associated with customers. When a user

begins entering customer information into a form, information associated with relevant existing customer records may be provided to the user such that the user may select a relevant record to automatically populate the form. The relevant records provided to the user may be based on the information entered by the user and may account for misspellings (e.g., dropped letters, insertions, replacements, transpositions, etc.), superfluous words, abbreviations, related words and/or phrase, aliases, acronyms, synonyms and/or equivalent terms, foreign languages, and the like, that may be associated with the information provided by the user. For example, if a user begins entering information for Customer A and a record associated with Customer A currently exists in the database, the user may select the Customer A record from a list of records provided to the user, where the list of records provided may include records that may be relevant to the information that the user has entered into the form. The list of records provided may be updated as the user types based on the relevancy of records with respect to the user input. In some examples, the list of records provided may account for one or more words that may be in the same field and/or different fields of the form.

[0008] When a user begins typing characters into a particular field of a form on a user interface, the query is received and normalized such that the characters of the query are transformed to a standardized format. For example, a character “Ã” in a query may be normalized to the character “a”. A mapping of the normalized query is generated for each record in a set of records in a database. The mapping is based on bit-shifting bits associated with similarities between each character in the normalized query and each character in a field of the respective record in the set of records, as described in more detail below. A score associated with each record is calculated based on the respective mapping, and the score is used to provide a subset of the set of records to the user in response to the query. For example, records within the set of records having higher scores may be provided to the user in response to the query.

[0009] Each record R_i in a database may consist of a list of one or more fields (e.g., name, street, city, state, postal code, country, etc.). In some

examples, the fields associated with a record may allow for synonym extensions to enable searching for a record based on synonyms for terms specified in the record. For example, if a company name has synonyms that include a pre-merger company name, the synonym extension may allow users to search for the company record using either name, as the record may include additional fields for synonyms like the pre-merger company name. Likewise, if the state field has an abbreviation alternative, it may be added as an additional field (e.g. "Alaska" and "AK").

[0010] A query may be normalized in any suitable manner such that the query may be standardized and compared with a normalized record. Examples of normalizations of queries and records include mapping characters to uppercase or lowercase (e.g., "A" → "a"), mapping accented characters to their unaccented version (e.g. Ü → U), mapping '&,' '+,' 'and,' or 'und' to '&' to accommodate for name variations entered in the database record such as "C and J's Plumbing", "C&J Plumbing", or "C + J Plumbing," mapping other punctuation to a space character (e.g., "," → " "), ensuring that multiple interior space characters are reduced to a single space, ensuring that a query string begins and ends with a single space character such that a query such as "IT" matches to the word "IT" in a record with a higher score than matching as a prefix to a word such as "Itanium" or to the interior of a word such as "Invitations," and the like. Queries and/or records may be normalized by leveraging an array that maps an input character to an output character. Records may be normalized in advance of comparing a record to a query or on-the-fly after a query is received.

[0011] In some examples, when database fields are normalized, a space character may be inserted before and after a sequence of digit characters if the digit characters are immediately preceded and/or followed by foreign language ideograph characters (e.g., Chinese characters, Japanese characters, Korean characters, etc.). For example, a database field that includes "浦东新区春耘路23号" may have space characters inserted before and after "23" such that a query having "23" may match the number "23" in the record with a higher score than records that only have this particular

sequence of digits buried within a longer number, such as “浦东新区春耘路1230号.”

[0012] In some examples, a query may consist of sixty-four characters or less, and thus, without loss of generality, the techniques described herein use examples of 64-bit long numbers. However, one having ordinary skill in the art will appreciate that longer queries may be processed similarly and/or by using additional processing, such as being divided and handled in pieces, being handled by providing a longer bit-wise representation, and the like. The techniques disclosed herein are described with the convention that the least-significant bit corresponds to the character at the beginning of a query, and thus a left-shift operator (e.g., $\ll$) may advance each bit to the next position in the query string.

[0013] Once a query has been received and normalized, a mapping that maps each character *ch* of the normalized query to all the bit positions that the character occupies may be generated. The mapped entry for a non-occurring character may return zero, indicating that the bit position is not enabled. In some examples, a `long[] charmasks` array may be used to quickly access the bit positions via a simple lookup `charmasks[ch]`.

[0014] A score may be calculated for each record R_i , where a score may indicate the degree to which a record corresponds to the received query. The scores for each record R_i may be calculated either sequentially or in parallel on multiple central processing unit (CPU) cores independently. To calculate a score of a record, an accumulator array of `query.length` long words may be allocated to the record (e.g., `accumulator := new long[qlen]`). Each position within a long may correspond to a position within the query string, and each successive array slot may correspond to more characters matched between the query and the record. The accumulator may record the higher matches over all text fields and over all positions within a field string. This may be considered as a high water mark for the state array *S*, which is described below. In some examples, the accumulator and state arrays may be shorter (e.g., `qlen` = the length of the longest term or phrase, including prefix and suffix spaces, that one may wish to be able to detect whole).

[0015] A state array S of query.length longs initially set to zero may be allocated for each field F of the record R (e.g., $S := \text{new long}[\text{qlen}]$), similar to the accumulator. Bits may be copied to successive array positions at successive bit positions when the next character in the query matches, and bits may be copied to antecedent array positions when a matching error is encountered, such as a character substitution (e.g., where the query and text disagree and each character forward is skipped), a query character deletion (e.g., the text character is skipped and the matching is continued from the same position), a query character insertion (e.g., the query character is skipped and matched at the following position), and the like.

[0016] For each character ch of field F , the character bit positions may be looked up from the mapping generated after receipt and normalization of the query (e.g., $\text{mask} := \text{charmasks}[ch]$), and a new state array S' may be computed. Since the new state array S' is an array of longs, the computation is iterated over each index position. For each $q = 0 \dots \text{qlen}-1$, the new state array S' is computed as follows:

- $\text{long more} := S[q+1]$, else 0 if $q+1$ is past the end of the array
- $\text{long same} := S[q]$
- $\text{long less} := S[q-1]$, else -1 (all 1 bits) if $q=0$
- $\text{long enabled} = (\text{less} | (\text{same} \ll 1)) \ll 1$, else -1 (all bits 1) if $q=0$, where the all-bits-enabled when $q=0$ is what allows the algorithm to begin a match at any character position at each step. The inclusion of " $\text{same} \ll 1$ " corresponds to a query deletion (e.g., -1 array position) followed by a query match (e.g., +1 array position).
- $S'[q] := (\text{mask} \& \text{enabled}) | \text{more} | (\text{more} \ll 1)$. Here, a bit is set to 1 at array position q if and only if the current character appears at a position right after a previously enabled position for position $q-1$ matched (e.g., $\text{mask} \& \text{enabled}$), or if and only if this position had a match for $q+1$ matched characters (e.g., more), or if and only if the previous position had a match for $q+1$ (e.g., $\text{more} \ll 1$).
- $\text{accumulator}[q] = \text{accumulator}[q] | S'[q]$. This keeps track of the highest array position enabled for each bit position.

- Optionally, if phrases between multiple words are not tracked, the bits in space positions for all $q > 0$ are zeroed out. In this case, $qlen$ need only be the length of the longest query term (e.g., if $(q > 0) S'[q] \&= \sim charmask[s']$; where $\sim$ is bitwise inversion).

[0017] Once the new state array S' is computed, state array S is set to new state array S' . This may be implemented with a pointer swap, and the old allocated array may be used for S' the next time.

[0018] Once all the fields in a record have been processed, the accumulator bits may be processed to compute the score for the record, where the initial score for each record may begin at a particular score (e.g., initial score of 0.0). For each query term end position i (e.g., each bit position corresponding to a space character, except the first), the maximum q value (e.g., q_{max}) is determined, where $accumulator[q]$ bit i is set or bit $i-1$ is set. This may find the maximum number of characters matching by the end of the term, with or without the final space. The credit for this term may then be computed. This may be done in a number of variations. The approximate percentage of matched characters may be computed (e.g., q_{max} divided by the length of the term including the space character on both ends). In some examples, the credit (e.g., in range from 0.0 to 1.0) may be squared to further penalize weak matches. If the optional step of zeroing out bits for phrases that are not tracked is not included, then the scoring of phrases may be handled by considering them worth more than their constituent terms alone. If q_{max} is greater than the current term length, then the database had text that covered a phrase of terms. In this case, the term length of the phrase is extended to cover additional previous terms until it is greater than or equal to q_{max} . The phrase credit is then q_{max} divided by the phrase length with 1.0 being a perfect match of the phrase including spaces on the ends, increased by a bonus for covering the phrase. There are a variety of ways to do this, such as increasing the phrase credit by an additional amount or percentage (e.g., +0.1, 25%, etc.) optionally increased for each term in the phrase. The score may then be calculated by adding the score and the credit, and the score may be returned for the entire record.

[0019] A subset of the records for which scores were determined may be presented to the user based on the scores for each record. For example, the ten highest-scoring records may be presented to the user in any suitable manner (e.g., as an auto-completion result, as a list of search hits, etc.).

[0020] In some examples, if there are too many records to search within the desired user-interactive latency bound, then a pre-filtering stage may be added to quickly select a subset of records on which to perform the bit-shift algorithm described above. The pre-filter may be a simple index of the set of n-grams of adjacent characters that occurs in the record anywhere (e.g., 3-grams, 2-grams, 1-grams, etc.). For example, when the query arrives, a set of 1- through 3-grams it contains may be produced, and the reverse indices may be used to tally how many of these n-grams are contained in each record. An array of numbers may be used to tally the count. A portion of the array (e.g., the 25,000 largest values in the array) may be used as candidate records for which to run through the bit-shift algorithm.

[0021] Referring now to the figures, FIG. 1 is an interface diagram of an example user interface 100 for accessing database records based on bit-shifting. User interface 100 may be used to enter customer information into a database. A user may enter information into the various fields of user interface 100 (e.g., company name, street, city, state, postal code, country, etc.). For example, a user may enter a query “anti square 3102” in the company name field 102. As the user types the information into field 102, drop-down menu 104 may appear, displaying existing records that may be relevant to the user’s entry in field 102, where the relevant records are identified using a mapping generated based on bit-shifting of bits associated with similarities between each character in the query and each character in a field of a record. For example, if a user begins typing customer information into field 102 for a particular customer, the user may use drop-down menu 104 to determine that the particular customer already has a relevant record in the database, and the user may select the relevant record entry (e.g., record entry 106) to automatically populate the fields in user interface 100, which may increase the accuracy and speed at which a user may find and enter customer information.

[0022] In some examples, a user may begin typing information in any field of user interface 100, and user interface 100 may display relevant records based on the typed information being relevant to any field of user interface 100. For example, a user may begin typing “anti square 3102” in field 102, and drop-down menu 104 may display records that contain similar text in any field of any record. As shown in the example of FIG. 1, drop-down menu 104 may display a record having “anti,” or a word similar to or a variation of “anti,” in the company name field 102 (e.g., Anti-Freeze Corporation, Antics, Antarctic Products, Ant Incorporated, Frantic Party Supply), street field (e.g., 432 Antithesis Square), city field, state field, postal code field, and/or country field.

[0023] In some examples, as the user types information into any field, the relevant records displayed in drop-down menu 104 may be updated based on newly typed information. For example, after the user types “anti” in field 102, drop-down menu 104 may display records relevant to “anti.” When the user types additional information in field 102, such as “square,” drop-down menu 104 may be updated with records relevant to the information typed in field 102 (e.g., “anti square” may retrieve some records that include “square” in the street field). A user may type into any field each portion of text to help distinguish the sought record, rather than having to carefully type each portion into the appropriate field. This may be helpful in cases where some database entries have data fields mixed up (e.g. the state and the city field swapped in the stored record).

[0024] FIG. 2 is a block diagram of an example computing device 200 for providing database records based on bit-shifting. Computing device 200 may receive a query, and in response, a subset of records in a database may be provided using a mapping generated based on bit-shifting of bits associated with similarities between each character in the query and each character in a field of each record.

[0025] Computing device 200 may be, for example, a web-based server, a local area network server, a cloud-based server, a notebook computer, a desktop computer, an all-in-one system, a tablet computing device, a mobile phone, an electronic book reader, a printing device, or any other electronic

device suitable for providing database records based on bit-shifting. Computing device 200 may include a processor 202 and a machine-readable storage medium 204. Computing device 200 may use a mapping generated based on bit-shifting of bits associated with similarities between each character in a query and each character in a field of a particular record in a database to provide a subset of records in response to the query.

[0026] Processor 202 is a tangible hardware component that may be a central processing unit (CPU), a semiconductor-based microprocessor, and/or other hardware devices suitable for retrieval and execution of instructions stored in machine-readable storage medium 204. Processor 202 may fetch, decode, and execute instructions 206, 208, 210, 212, and 214 to control a process of providing database records based on bit-shifting. As an alternative or in addition to retrieving and executing instructions, processor 202 may include at least one electronic circuit that includes electronic components for performing the functionality of instructions 206, 208, 210, 212, 214, or a combination thereof.

[0027] Machine-readable storage medium 204 may be any electronic, magnetic, optical, or other physical storage device that contains or stores executable instructions. Thus, machine-readable storage medium 204 may be, for example, Random Access Memory (RAM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), a storage device, an optical disc, and the like. In some examples, machine-readable storage medium 204 may be a non-transitory storage medium, where the term “non-transitory” does not encompass transitory propagating signals. As described in detail below, machine-readable storage medium 204 may be encoded with a series of processor executable instructions 206, 208, 210, 212, and 214 for receiving a query; creating a normalized query that is a normalized version of the query; generating a mapping of the normalized query for each record in a set of records in a database using a state array, the mapping being based on bit-shifting of bits associated with similarities between each character in the normalized query and each character in a field of the respective record in the set of records, each successive level of the state array indicating a stronger match between the query and the respective record than an immediately prior

level in the state array; calculating a score associated with each record based on the respective mapping; and providing a subset of the set of records in response to the query based on the score associated with each record in the subset of records.

[0028] Query receipt instructions 206 may manage and control receipt of queries from users, including managing and controlling the receipt of characters in the queries. For example, query receipt instructions 206 may receive one or more characters in a query inputted into a field in user interface 100 of FIG. 1.

[0029] Normalization instructions 208 may manage and control the normalization of a query. For example, normalization instructions 208 may normalize the received query such that each character of the query is transformed into a standardized format (e.g., normalizing “Ã” to “a”), creating a normalized query that is a normalized version of the query. In some examples, normalization instructions 208 may also normalize each record in a set of records in a database into a normalized record. For example, normalization instructions 208 may normalize a set of pre-filtered records to create a set of normalized pre-filtered records.

[0030] Map generation instructions 210 may manage and control the generation and/or creation of a mapping of the normalized query for each record in a set of records in a database using a state array. The mapping may be generated and/or created based on bit-shifting of bits associated with similarities between each character in the normalized query and each character in a field of the respective normalized record in the set of records. In some examples, the mapping may be generated using a state array, where each successive level of the state array indicates a stronger match between the query and the respective record than an immediately prior level in the state array. In some examples, the mapping may be generated based on an enabled bit associated with a query character being shifted if the query character is similar to a corresponding record character. In some examples, the enabled bit is not shifted if a subsequent query character is different than a corresponding record character.

[0031] Score calculation instructions 212 may manage and control the determination and/or calculation of a score associated with each record based on the respective mapping for that record. For example, a particular record may be scored based on the mapping generated based on bit-shifting of bits (e.g., based on the similarities between each character in the normalized query and each character in a field of the respective normalized record). In some examples, the scores associated with each record may be tracked using an accumulator array.

[0032] Record retrieval instructions 214 may manage and control the retrieval and the providing and/or displaying of a subset of the set of records in response to the query. The subset of records may be retrieved and provided and/or displayed based on the score associated with each record in the subset of records. In some examples, the subset of records may be records having scores within a particular range with respect to scores of other records in the set of records (e.g., providing records that have a score within the top 10% of scores for the set of records). In some examples, the subset of records may be displayed in an order that is based on the scores of each record (e.g., the higher scoring records being displayed at the top of the list of records provided).

[0033] FIG. 3 is a flowchart illustrating an example method 300 of returning database records based on bit-shifting. Method 300 may be implemented using computing device 200 of FIG. 2.

[0034] Method 300 includes, at 302, receiving a query. For example, in FIG. 1, the query “anti square 3102” in field 102 may be received via user interface 100.

[0035] Method 300 also includes, at 304, standardizing the query into a normalized query. For example, a query containing capital letters, accented letters, punctuation, and the like may be standardized to a normalized query.

[0036] Method 300 also includes, at 306, creating a normalized record for each record in a set of records in a database. For example, a record in a database may be normalized to create a normalized record by removing

additional elements of each character in the record, such as capital letters, accented letters, punctuation, and the like.

[0037] Method 300 also includes, at 308, producing a mapping of the normalized query for each normalized record using a state array. The mapping may be based on bit-shifting of bits associated with similarities between each character in the normalized query and each character in a field of the respective normalized record. In some examples, the mapping may be produced using a state array, where each successive level of the state array indicates a stronger match between the query and the respective normalized record than an immediately prior level in the state array.

[0038] Method 300 also includes, at 310, identifying a score associated with each record based on the respective mapping. In some examples the score for each record may be tracked using an accumulator array.

[0039] Method 300 also includes, at 312, returning a subset of the set of records in response to the query based on the score associated with each record in the subset of records. In some examples, the subset of records may include records that fall within a particular range of scores associated with records in the database. In some examples, the subset of records may be returned in an order that is based on the score associated with each record.

[0040] Examples provided herein (e.g., methods) may be implemented in hardware, software, or a combination of both. Example systems may include a controller/processor and memory resources for executing instructions stored in a tangible non-transitory medium (e.g., volatile memory, non-volatile memory, and/or machine-readable media). Non-transitory machine-readable media can be tangible and have machine-readable instructions stored thereon that are executable by a processor to implement examples according to the present disclosure.

[0041] An example system can include and/or receive a tangible non-transitory machine-readable medium storing a set of machine-readable instructions (e.g., software). As used herein, the controller/processor can include one or a plurality of processors such as in a parallel processing system. The memory can include memory addressable by the processor for

execution of machine-readable instructions. The machine-readable medium can include volatile and/or non-volatile memory such as a random access memory ("RAM"), magnetic memory such as a hard disk, floppy disk, and/or tape memory, a solid state drive ("SSD"), flash memory, phase change memory, and the like.

Claims

What is claimed is:

1. A computing device comprising:
at least one processor to:
receive a query;
create a normalized query, the normalized query being a normalized version of the query;
generate a mapping of the normalized query for each record in a set of records in a database using a state array, the mapping being based on bit-shifting of bits associated with similarities between each character in the normalized query and each character in a field of the respective record in the set of records, each successive level of the state array indicating a stronger match between the query and the respective record than an immediately prior level in the state array;
calculate a score associated with each record based on the respective mapping; and
provide a subset of the set of records in response to the query based on the score associated with each record in the subset of records.
2. The computing device of claim 1, wherein the at least one processor is further to pre-filter records in the database based on the query, the set of records being based on the pre-filtering.
3. The computing device of claim 1, wherein the subset of records includes records having scores within a particular range with respect to scores of other records in the set of records.
4. The computing device of claim 1, wherein the mapping is generated based on an enabled bit associated with a query character being shifted if the query character is similar to a corresponding record character.

5. The computing device of claim 1, wherein an array length of the state array is based on a query length of the query.

6. The computing device of claim 1, wherein the score associated with each record is tracked using an accumulator array and wherein the accumulator array includes a zero associated with each space character in the field of the respective record.

7. A method comprising:
receiving, by a computing device, a query;
standardizing, by the computing device, the query into a normalized query;
creating, by the computing device, a normalized record for each record in a set of records in a database;
producing, by the computing device, a mapping of the normalized query for each normalized record using a state array, the mapping being based on bit-shifting of bits associated with similarities between each character in the normalized query and each character in a field of the respective normalized record, each successive level of the state array indicating a stronger match between the query and the respective normalized record than an immediately prior level in the state array;
identifying, by the computing device, a score associated with each record based on the respective mapping; and
returning, by the computing device, a subset of the set of records in response to the query based on the score associated with each record in the subset of records.

8. The method of claim 7, further comprising:
pre-filtering records in the database based on the query, the set of records being based on the pre-filtering.

9. The method of claim 7, wherein an array length of the state array is based on a query length of the query.

10. The method of claim 7, wherein the mapping is produced based on an enabled bit associated with a query character being shifted if the query character is similar to a corresponding record character.

11. A non-transitory machine-readable storage medium storing instructions that, if executed by at least one processor of a computing device, cause the computing device to:

- receive a query;

- normalize the query into a normalized query;

- normalize each record in a set of records in a database into a normalized record;

- create a mapping of the normalized query for each normalized record using a state array, the mapping being based on bit-shifting of bits associated with similarities between each character in the normalized query and each character in a field of the respective normalized record, each successive level of the state array indicating a stronger match between the query and the respective normalized record than an immediately prior level in the state array;

- determine a score associated with each record based on the respective mapping; and

- display a subset of the set of records in response to the query based on the score associated with each record in the subset of records.

12. The non-transitory machine-readable storage medium of claim 11, wherein the instructions further cause the computing device to:

- pre-filter records in the database based on the query, the set of records being based on the pre-filtering.

13. The non-transitory machine-readable storage medium of claim 11, wherein an array length of the state array is based on a query length of the query.

14. The non-transitory machine-readable storage medium of claim 11, wherein the mapping is created based on an enabled bit associated with a query character being shifted if the query character is similar to a corresponding record character.

15. The non-transitory machine-readable storage medium of claim 11, wherein the score associated with each record is tracked using an accumulator array and wherein the accumulator array includes a zero associated with each space character in the field of the respective record.

100 →

CUSTOMER INFORMATION

Company:102anti square 3102

Street:104ANTI-FREEZE CORPORATION; 123 MAIN STREET; SAN

City:106ABC MARKET; 432 ANTITHESIS SQUARE; SAN JOSE; C

State:ANTICS; 90 GOLCHA SQUARE; OAKLAND; CALIFORNIA

State:ANTARCTIC PRODUCTS; 310 2ND AVENUE; PALO ALTO

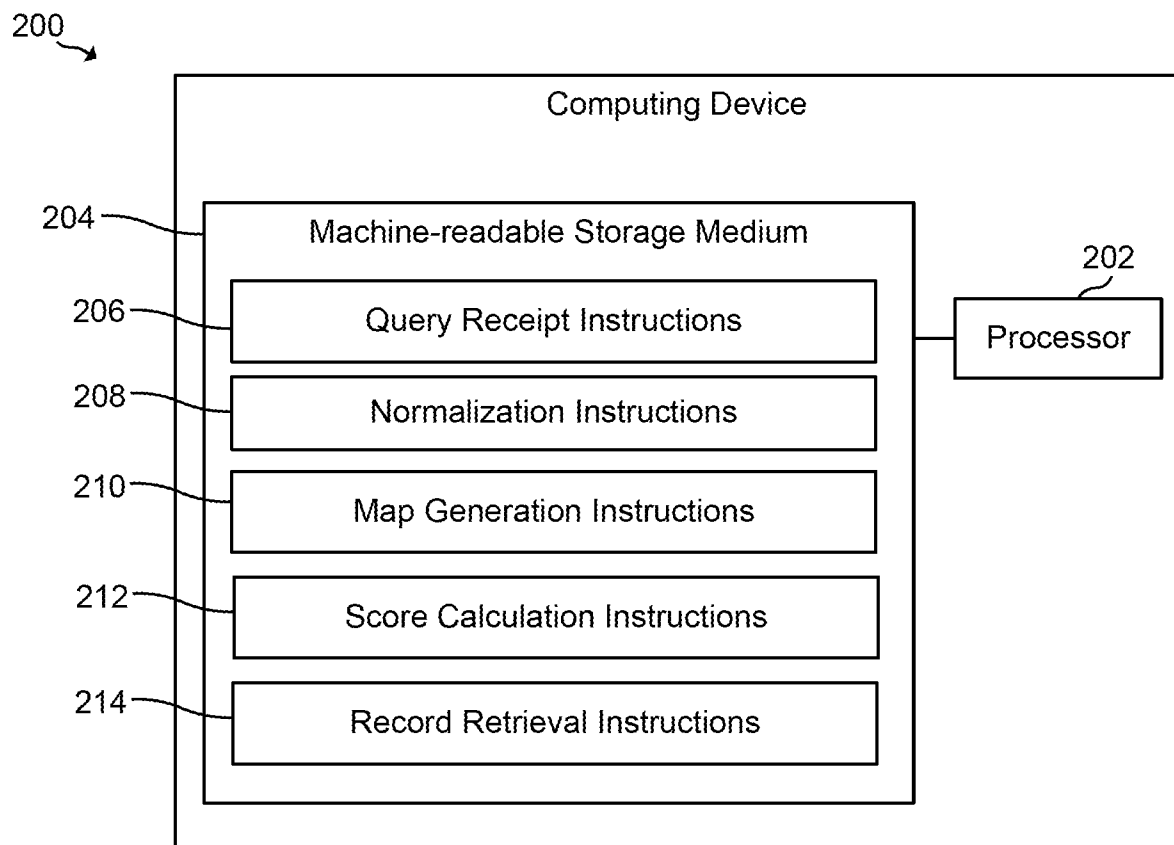
State:ANT INCORPORATED; 32 UNION SQUARE; MENLO PARK

Postal Code:FRANTIC PARTY SUPPLY; 102 MARKET STREET; SQUAW

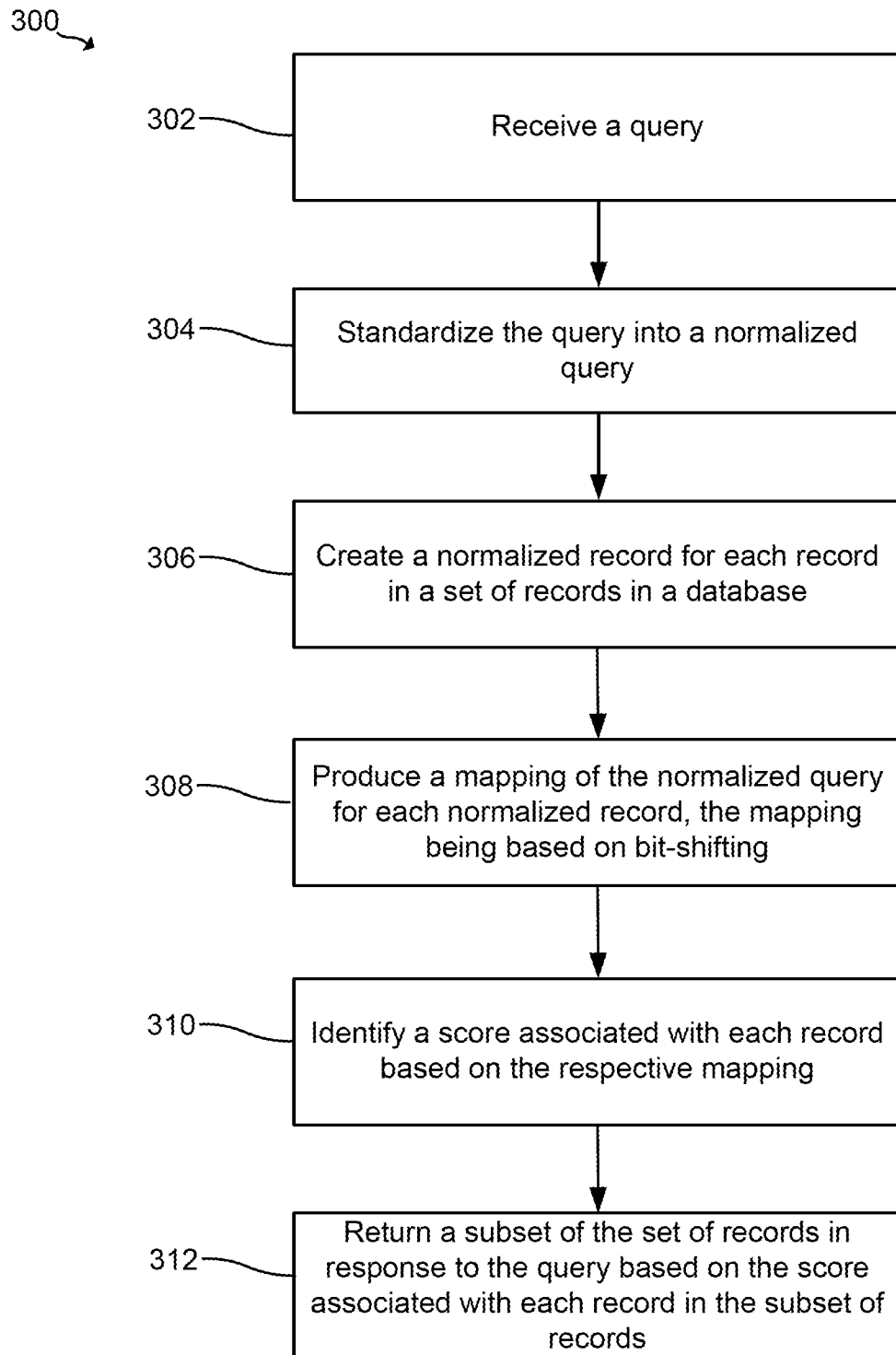
Country:

FIG. 1

2/3

**FIG. 2**

3/3

**FIG. 3**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/014300**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/30(2006.01)i**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/30; G06F 7/00; G06F 12/00; G09G 5/00; G06F 7/06

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: query, normalize, database record, match, character, mapping, bit-shifting, successive level, score, and similar terms.**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013-0097157 A1 (PO CHEUNG NG et al.) 18 April 2013 See paragraphs [0019], [0022], and [0025]-[0026]; claim 1; and figures 1-2.	1-15
A	US 7,596,553 B2 (ORNA MEIROVITZ et al.) 29 September 2009 See column 4, lines 39-54; column 4, line 64 - column 5, line 5; column 5, lines 38-49; column 6, lines 4-13; column 6, line 36 - column 7, line 32; claims 20 and 32-33; and figures 1-3A.	1-15
A	US 7,155,427 B1 (GAUTAM PROTHIA et al.) 26 December 2006 See column 7, line 9 - column 8, line 37; claim 1; and figures 1-2.	1-15
A	US 2008-0319990 A1 (ALEXANDRE TARANENKO et al.) 25 December 2008 See paragraphs [0072]-[0076] and figures 12-14.	1-15
A	US 8,463,810 B1 (EARL RENNISON) 11 June 2013 See column 27, line 25 - column 29, line 2; claim 1; and figure 2.	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of 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

30 November 2015 (30.11.2015)

Date of mailing of the international search report

30 November 2015 (30.11.2015)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 35208,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

NHO, Ji Myong

Telephone No. +82-42-481-8528



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/014300

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2013-0097157 A1	18/04/2013	CA 2792070 A1 MX 2012011903 A	14/04/2013 07/08/2013
US 7596553 B2	29/09/2009	US 2004-0073550 A1	15/04/2004
US 7155427 B1	26/12/2006	None	
US 2008-0319990 A1	25/12/2008	EP 2158540 A1 EP 2158540 A4 US 8015196 B2 WO 2008-156600 A1	03/03/2010 20/10/2010 06/09/2011 24/12/2008
US 8463810 B1	11/06/2013	US 7827125 B1 US 7870117 B1 US 8001064 B1 US 8024329 B1 US 8620909 B1	02/11/2010 11/01/2011 16/08/2011 20/09/2011 31/12/2013