

METHOD AND APPARATUS FOR MATCHING NAMES

TECHNICAL FIELD

[0001] The present application relates to the field of computer software
5 technologies, and in particular, to a method and an apparatus for matching names.

BACKGROUND

[0002] Person name matching is a very important technology in the risk control
field. For example, a risk control system records person names of determined
unauthorized users in a blacklist. Then, when performing risk control operation, for
10 each user that currently performs a service, a person name of each user is matched
with each person name in the blacklist through scanning. If the matching succeeds,
the user can be considered as an unauthorized user, and the service of the user is
rejected, to prevent certain risks.

[0003] Person name matching can be classified into accurate person name
15 matching and person name fuzzy matching. In comparison, person name fuzzy
matching is more difficult in terms of technologies because it is difficult to control a
proper fuzzy degree.

[0004] In the existing technology, a string matching algorithm is usually used to
perform person name fuzzy matching, and a string matching degree threshold
20 determines a fuzzy degree. However, the string matching degree threshold is all set
according to experience. To reduce omission, the string matching degree threshold is
usually set to a relatively low value. Consequently, the matching accuracy is relatively
low, and a false alarm rate of the risk control system is relatively high.

SUMMARY

25 [0005] Implementations of the present application provide a method and an
apparatus for matching names, to alleviate the following technical problem: In the

existing technology, the relatively low matching accuracy and relatively high system false alarm rate happen when person name fuzzy matching is performed by using a string matching algorithm.

[0006] To alleviate the previous technical problem, the implementations of the present application are implemented as follows:

[0007] An implementation of the present application provides a method for matching names, including the following: obtaining a name to be matched; determining a standard name set used to match the name to be matched; detecting the name to be matched to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical; and determining a matching result of the name to be matched based on a detection result.

[0008] An implementation of the present application provides an apparatus for matching names, including the following: an acquisition module, configured to obtain a name to be matched; a determining module, configured to determine a standard name set used to match the name to be matched; a detection module, configured to detect the name to be matched to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical; and a matching module, configured to determine a matching result of the name to be matched based on a detection result.

[0009] The at least one technical solution used in the implementations of the present application can achieve the following beneficial effects: The name can include a person name. In practice, a person name to be matched may be different from an actual person name due to the timeliness and the uncertainty during data entry and the variability of the person name. This is also the reason for performing fuzzy matching. In the solution of the present application, for this reason, the person name to be matched is detected to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical, and a person name matching result is determined based on a detection result. Compared with the existing technology that a fuzzy degree is controlled only by using a string matching degree threshold set according to experience, the present application is more conducive to improving the accuracy of controlling the fuzzy degree. As such, the matching accuracy can be improved, and a false alarm rate of a risk control system can be reduced. Therefore, some or all of the problems in the existing technologies

can be alleviated.

BRIEF DESCRIPTION OF DRAWINGS

[0010] To describe the technical solutions in the implementations of the present application or in the existing technology more clearly, the following briefly describes the accompanying drawings required for describing the implementations or the existing technology. Apparently, the accompanying drawings in the following description merely show some implementations of the present application, and a person of ordinary skill in the art can still derive other drawings from these accompanying drawings without creative efforts.

[0011] FIG. 1 is a schematic flowchart illustrating a method for matching names, according to an implementation of the present application;

[0012] FIG. 2 is a schematic flowchart illustrating an implementation of preliminary screening in a method for matching names in an actual application scenario, according to an implementation of the present application;

[0013] FIG. 3 is a schematic flowchart illustrating an implementation of performing fuzzy matching by using an integrated algorithm in a method for matching names in an actual application scenario, according to an implementation of the present application;

[0014] FIG. 4 is a schematic flowchart illustrating an implementation of a method for matching names in an actual application scenario, according to an implementation of the present application; and

[0015] FIG. 5 is a schematic structural diagram illustrating an apparatus for matching names and corresponding to FIG. 1, according to an implementation of the present application.

DESCRIPTION OF IMPLEMENTATIONS

[0016] Implementations of the present application provide a method and an apparatus for matching names.

[0017] To make a person skilled in the art understand the technical solutions in the present application better, the following clearly and comprehensively describes the technical solutions in the implementations of the present application with reference to

the accompanying drawings in the implementations of the present application. Apparently, the described implementations are merely some but not all of the implementations of the present application. All other implementations obtained by a person of ordinary skill in the art based on the implementations of the present application without creative efforts shall fall within the protection scope of the present application.

[0018] As described above, in practice, a person name to be matched may be different from an actual person name due to the timeliness and the uncertainty during data entry and the variability of the person name (which mainly indicating a change in a "shape (namely, character)" of the person name). For ease of understanding, an English person name is used as an example. Common variation types and instances of an English person name are shown in Table 1.

Table 1

Common variation type of an English person name	Instance of a person name variation type
Misspelling	Jaxson, Jakson, and Jaxon are originally expected to represent the same person name. SMYTH and SMITH are originally expected to represent the same person name. Spelling error.
Phonetic spelling difference	Michel, Michal, and Miguel are originally expected to represent the same person name. Misspelling caused by similar pronunciations.
Nickname	Mike, Mick, and Mikey represent nicknames of the same person.
Address term	Dr., Mr., etc. are terms of addressing a person, and do not affect a meaning of a person name when they appear together with the person name.
Synonym	Mohamed, Mohammed, and Mohammad are synonymous when appearing in a person name.
Abbreviation	A person name ROB is an abbreviation for ROBBIN, and ROB and ROBBIN are synonymous. A person name BOB is an abbreviation for BOBBY, and ROB and BOBBY are synonymous.
Multi-national language difference	Russian BORISOV has a different meaning from BORISOVNA, and Spanish JUANLTA is synonymous with JUAN.

[0019] If only a string matching method is used to match a person name, it is possible that matching can be attempted for a variation type of "misspelling" a certain letter (actually, the matching succeeds accidentally, and there is no solid basis).

5 However, the matching accuracy is very low for other variation types.

[0020] In the solutions of the present application, particular detection such as abbreviation detection, address term detection, multi-language detection, or alias detection can be performed for the other variation types, so that a situation that person names are synonymous but characters of the person names are not identical (in other words, the names are synonymous but different in shapes) can be considered comprehensively, and further the matching accuracy can be improved. It is worthwhile to note that "different shapes" can be different shapes that are "wrong" due to misspelling; but in the following implementations, can be mainly different shapes that are "reasonable and correct" due to the other variation types.

15 [0021] The solutions of the present application are not only applicable to person name matching, but also applicable to matching of names other than a person name, for example, a place name or an object name.

[0022] The solutions of the present application are described below in detail.

[0023] FIG. 1 is a schematic flowchart illustrating a method for matching names, according to an implementation of the present application. A device that a program executing the procedure can be installed on includes but is not limited to a personal computer, a large or medium computer, a computer cluster, a mobile phone, a tablet computer, a smart wearable device, a vehicle machine, etc. This process usually can be used in the risk control field, and is executed by a risk control system or a related system.

25 [0024] The procedure shown in FIG. 1 can include the following steps:

[0025] S101. Obtain a name to be matched.

[0026] In the present implementation of the present application, a language that the name to be matched belongs to is not limited, and can be English, Russian, Spanish, or Chinese. For ease of description, the following implementation is mainly described by using an example that the language that the name to be matched belongs to is English.

[0027] S102. Determine a standard name set used to match the name to be matched.

[0028] In the present implementation of the present application, the standard name set can be a subset screened out from a larger name set, or can be directly the larger name set. For ease of description, the screening here can be referred to as "preliminary screening". For example, in a scenario in the background, the larger name set can be a blacklist held by a risk control system.

[0029] In the previous situation, the subset can be a set that includes only names similar to the name to be matched. "Similar" here may not be so strict because there is a series of subsequent operations of further determining the similarity. It means that "fine screening" is performed subsequently. A matching range can be quickly reduced through preliminary screening, so that a workload of subsequent fine screening can be reduced, and the pertinence of fine screening can be improved, which is conducive to improving the efficiency of the solutions of the present application.

[0030] S103. Detect the name to be matched to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

[0031] In the present implementation of the present application, "the names are synonymous but characters of the names are not identical" that needs to be detected and determined is mainly caused by one or more variation types in Table 1, and the detection can include at least one of detections such as abbreviation detection, address term detection, multi-language detection, or alias detection, which are described below in detail.

[0032] In the present implementation of the present application, when there are a plurality of detections, the plurality of detections can be successively performed in order. Remaining detections may not be performed if a matching result of the name to be matched can be determined in the detection process. Certainly, to improve the execution efficiency, the plurality of detections can be performed in parallel and then detection results can be summarized.

[0033] S104. Determine a matching result of the name to be matched based on a detection result.

[0034] In the present implementation of the present application, by performing step S103, the matching result of the name to be matched can be directly determined if it is determined that the name to be matched is synonymous with at least one name

in the standard name set but characters of the names are not identical. In this situation, the detection process in step S103 is actually an entire matching process of the name to be matched.

[0035] If it is determined that the name to be matched does not satisfy the following condition: the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical, matching can be further performed on the name to be matched by using another matching method to determine the matching result of the name to be matched.

[0036] Through the method shown in FIG. 1, the name can include a person name. In practice, a person name to be matched may be different from an actual person name due to the timeliness and the uncertainty during data entry and the variability of the person name. This is also the reason for performing fuzzy matching. In the solution of the present application, for this reason, the person name to be matched is detected to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical, and a person name matching result is determined based on a detection result. Compared with the existing technology that a fuzzy degree is controlled only by using a string matching degree threshold set according to experience, the present application is more conducive to improving the accuracy of controlling the fuzzy degree. As such, the matching accuracy can be improved, and a false alarm rate of the risk control system can be reduced. Therefore, some or all of the problems in the existing technologies can be alleviated.

[0037] Based on the method in FIG. 1, the present implementation of the present application further provides some implementations of the method and an extension solution, which are described below.

[0038] In the present implementation of the present application, the complexity of different names to be matched can be different, and included information can also be different. For some names to be matched that have a small volume of information or a simple information feature, it is difficult for the value of an obtained matching result to reach an expectation even if matching is performed on the names to be matched. For example, if a name is too simple and general, such as an English person name "Jim", "Jimmy", "David", "John", or "Mike", it is difficult to determine a certain person even if matching succeeds.

[0039] To alleviate the waste of processing resources used for name matching due

to this situation, the name to be matched can be first filtered after being obtained, to determine whether to continue the matching. In addition, if the name to be matched is included in a white list, the similar problem exists, and the name to be matched can also be processed by using this method.

5 **[0040]** For steps S101 and S102, after the name to be matched is obtained, and prior to the standard name set used to match the name to be matched is determined, the following steps can be further performed: obtaining a predetermined set of names that do not need to be matched; determining whether the name to be matched is included in the set of names that do not need to be matched; and if yes, continuing to
10 perform subsequent steps. Otherwise, matching may not be performed on the name to be matched.

[0041] In the present implementation of the present application, an implementation of step S102 (which corresponds to the previous situation of "preliminary screening") is used as an example for description. For step S102, the
15 determining a standard name set used to match the name to be matched can include the following: determining a first name set that can be used to match the name to be matched; and determining the standard name set used to match the name to be matched by performing similarity matching on each word included in the name to be matched and each word included in a name in the first name set.

20 **[0042]** There are also a plurality of implementations of how to perform similarity matching. Word segmentation matching can be performed on the name to be matched, or full-text matching can be performed on the name to be matched, etc.

[0043] In an example of word segmentation matching, the determining the standard name set used to match the name to be matched by performing similarity
25 matching on each word included in the name to be matched and each word included in a name in the first name set can include the following: obtaining an index of each name included in the first name set, where the index of the name is any word included in the name; segmenting the name to be matched to obtain each word included in the name to be matched; performing similarity matching on each word included in the
30 name to be matched and each index to obtain a subset of the first name set. The obtained subset includes a name indexed by each index that is successfully matched; and determining the standard name set used to match the name to be matched based on each subset.

[0044] The index in the previous example is pre-established, and an advantage of

performing word segmentation matching based on the index is that a speed of obtaining a name needed in the set in the matching process can be effectively accelerated. If word segmentation matching is not performed based on the index, word segmentation matching still can be implemented (for example, a data table that stores the set is directly queried for a needed name by using a Select statement for word segmentation matching), except that the efficiency may be affected.

[0045] Further, the performing similarity matching on each word included in the name to be matched and each index can include the following: performing similarity matching on each word included in the name to be matched and each index by using a string matching algorithm, where the string matching algorithm can include one or more algorithms, for example, a prefix tree matching algorithm, a dictionary tree matching algorithm, a string similarity matching algorithm, and a pronunciation similarity matching algorithm. Here, using the string matching algorithm is merely a better method, or another algorithm, such as a text matching algorithm, that can be used to implement similarity matching can be used.

[0046] There may be a plurality of implementations of determining the standard name set used to match the name to be matched based on each subset. For example, if N string matching algorithms are used, for each of M words obtained through word segmentation, each string matching algorithm is used to match the word with each index to correspondingly obtain N subsets. Then a union set of the N subsets is taken. A total of M union sets are obtained, and each name whose total number of occurrences in the M union sets exceeds a specified threshold is selected and determined as the standard name set. For another example, N subsets are obtained, and then an intersection set of the N subsets is taken and determined as the standard name set.

[0047] The "preliminary screening" process is described above. For ease of understanding, an implementation of the present application further provides a schematic flowchart illustrating an implementation of the preliminary screening in the method for matching names in an actual application scenario, as shown in FIG. 2.

[0048] In the actual application scenario, assume that the first name set is a list of English person names, an index of each name in the list is predetermined (related information other than the corresponding name can be indexed by using the index), the index is a word in the name that corresponds to the index, each index is included in an index table established by using a word that corresponds to the index as a

primary key, the name to be matched is an English person name "Kit Wai Jackson Wong", and a space is used as a delimiter to segment the person name, and a word segmentation result is shown in Table 2.

Table 2

Person name to be matched	Kit Wai Jackson Wong			
After word segmentation	Word 1: Kit	Word 2: Wai	Word 3: Jackson	Word 4: Wong

5

[0049] In FIG. 2, $M=4$, and the word segmentation result is {Kit, Wai, Jackson, Wong}. Each word obtained through word segmentation is matched with each index by using a prefix tree matching algorithm, a dictionary tree matching algorithm, a string similarity matching algorithm (such as a Simstring algorithm), and a pronunciation similarity matching algorithm (such as a metaphone algorithm), to output a name set that corresponds to indexes of a single word that are obtained by using the four matching algorithms: matching results 1, 2, 3, and 4.

10

[0050] A union set of the matching results of each word is taken to obtain a comprehensive matching result of each word.

15

[0051] Names included in at least two comprehensive matching results are selected to form a set, and the set is used as a generated preliminary screening result.

[0052] In the present implementation of the present application, after the standard name set is determined, the name to be matched can be detected. However, one or more types of preprocessing can be further performed prior to the detection. The preprocessing is conducive to improving the reliability of a subsequent detection result. The preprocessing can include alignment processing, capital and lowercase unification processing, simplified-traditional processing, etc.

20

[0053] Alignment processing is used as an example. For step S103, prior to the detecting the name to be matched, the following processing can be further performed: aligning each word included in a name in the standard name set with each word included in the name to be matched based on the similarity between each word included in the name in the standard name set and each word included in the name to be matched.

25

[0054] In practice, a principle of performing alignment processing based on the similarity can be a similarity-based maximization principle. To be specific, locations of names that include words with the maximum similarity are aligned.

[0055] For example, assume that the name to be matched is "Kate Lee Smith",
 5 and an alignment result of a name such as "Smith Catherine Lee" in the standard name set is shown in Table 3.

Table 3

Name to be matched	Kate	Lee	Smith
Name obtained after alignment processing in the standard name set	Catherine	Lee	Smith

[0056] Further, when alignment processing is performed, for step S103, the
 10 detecting the name to be matched can include the following: detecting the name to be matched based on the aligned standard name set to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

[0057] In the present implementation of the present application, the detection in
 15 step S103 is a key part for improving the matching accuracy of person name fuzzy matching. For step S103, the detecting the name to be matched includes the following: performing at least one of abbreviation detection, address term detection, multi-language detection, or alias detection on the name to be matched to determine whether the name to be matched is synonymous with at least one name in the standard
 20 name set but characters of the names are not identical. The several detections are separately described below.

[0058] In the present implementation of the present application, acronym
 detection is most common in the abbreviation detection. In addition to the acronym detection, there is abbreviation detection with omission of a partitive. In an
 25 implementation, the performing abbreviation detection on the name to be matched can include the following: obtaining predetermined abbreviation contrast combination data, where each abbreviation contrast combination reflects an abbreviation mapping relationship between at least one word and an abbreviation of the word; detecting whether a word included in the name to be matched has the abbreviation mapping

relationship with a word included in a name in the standard name set based on the abbreviation contrast combination data; and determining, based on a detection result, whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

5 **[0059]** For example, if it is predetermined that an English person name "Ben Williams" can be abbreviated to "B. Williams", "Ben Williams" and "B. Williams" can be used as an abbreviation contrast combination. If it is detected that the name to be matched and a name in the standard name set are the abbreviation contrast combination, it can be determined that there is a situation that the names are
10 synonymous but characters of the names are not identical.

[0060] In the present implementation of the present application, the performing address term detection on the name to be matched can include the following: obtaining predetermined address term data; detecting whether the name to be matched includes the address term, where it is considered that the address term does not affect
15 a meaning of the name that includes the address term; and determining, based on a detection result, whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

[0061] In terms of a person name, an address term is generally a word attached to at least a part of the original person name, such as an honorific title (such as Mr. or
20 Miss.) or a title (such as Dr. or Prof.). In a person name matching environment, the address term does not affect a meaning of the at least a part of the original person name that corresponds to the address term. Therefore, if it is determined that the name to be matched includes the address term, and other parts other than the address term can match a name in the standard name set, it can be determined that the name to be
25 matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

[0062] In the present implementation of the present application, the performing alias detection on the name to be matched can include the following: obtaining predetermined alias contrast combination data, where each alias contrast combination
30 reflects an alias mapping relationship between at least one word and an alias of the word; detecting whether a word included in the name to be matched has the alias mapping relationship with a word included in a name in the standard name set based on the alias contrast combination data; and determining, based on a detection result, whether the name to be matched is synonymous with at least one name in the standard

name set but characters of the names are not identical.

[0063] In practice, the alias can include a nickname (for example, Mick in Table 1 is a nickname of Mikey) of a name that corresponds to the alias or a synonymous name of the name that corresponds to the alias in different fields. For the latter, different fields can be different regions (for example, different countries or different provinces), different languages (for example, languages of different countries or languages of different nationalities), different industries, etc.

[0064] Correspondingly, at least one of the following is performed: a nickname of the name to be matched is detected, or a synonymous name of the name to be matched in different fields is detected.

[0065] The alias also does not affect a meaning of the name that corresponds to the alias. Therefore, if it is determined that the name to be matched is an alias of a name in the standard name set, it can be determined that the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

[0066] It is worthwhile to note that storage forms of the contrast combination data and the address term data described above are not limited in the present application. A common method is to store the contrast combination data and the address term data in a corresponding data table and read the data from a database when the data needs to be used.

[0067] In the present implementation of the present application, the performing multi-language detection on the name to be matched can include the following: determining a language that corresponds to the name to be matched; obtaining at least one of a spelling deformation synonym rule or a spelling deformation homonym rule of at least one of the language or another language; and detecting the name to be matched according to at least one of the spelling deformation synonym rule or the spelling deformation homonym rule to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

[0068] Multi-language detection is mainly aimed at the following situation, for example, English "Pooh" is spelled as "puh" in German, and the two words are synonymous when appearing in person names.

[0069] It is worthwhile to note that an algorithm used to calculate a string matching degree can be used in the detections performed previously.

[0070] In the present implementation of the present application, the determining a matching result of the name to be matched based on a detection result can include the following: determining the at least one name as the matching result of the name to be matched if it is determined that the name to be matched is synonymous with the at least one name in the standard name set but the characters of the names are not identical; or otherwise (to be specific, when the matching fails through the previous detection), determining the matching result of the name to be matched by matching the name to be matched with a name in the standard name set by using one or more similarity algorithms.

[0071] In the present implementation of the present application, the plurality of similarity algorithms can be based on different dimensions, and therefore can be conducive to improving the reliability of the matching result. According to such an idea, the similarity algorithm can be an algorithm (such as an n-gram algorithm) used to calculate a text matching degree, an algorithm (such as a Phonex algorithm) used to calculate a phonetic matching degree, an algorithm (such as a Jaro-Winkler algorithm) used to calculate a string matching degree, etc.

[0072] When the plurality of similarity algorithms are used, matching results that correspond to the similarity algorithms can be comprehensively measured to obtain a comprehensive matching result. A measurement method is not limited in the present application, and a common method is weighting summation.

[0073] For example, when the n-gram algorithm is used, an algorithm input is each word included in the name to be matched and each word at an aligned location of the word, and an algorithm output is a text matching degree of each pair of aligned words, and is denoted as F1.

[0074] When the Jaro-Winkler algorithm is used, an algorithm input is each word included in the name to be matched and each word at an aligned location of the word, and an algorithm output is a string matching degree of each pair of aligned words, and is denoted as F2.

[0075] When the Phonex algorithm is used, an algorithm input is each word included in the name to be matched and each word at an aligned location of the word, and an algorithm output is a phonetic matching degree of each pair of aligned words, and is denoted as F3.

[0076] A comprehensive matching degree F of each pair of aligned words is obtained by performing weighting summation on the text matching degree, the string

matching degree, and the phonetic matching degree, as shown in the following equation:

$$F=w_1*F_1+w_2*F_2+w_3*F_3, \text{ where } w_1+w_2+w_3=1.$$

[0077] Based on the comprehensive matching degree F of each pair of words, a result of matching between the name to be matched and the name in the standard name set is obtained by calculating an average value. For example, for a pair of names in Table 3, an obtained matching result is shown in Table 4.

Table 4

Detected name	Kate	Lee	Smith
Name in a list	Catherine	Lee	Smith
Comprehensive matching degree	0.872	1	1
Matching result	$(0.872+1+1)/3=0.957$		

[0078] The detection and matching process after the preliminary screening is described above. In this process, a plurality of matching-related algorithms can be used. During implementation of the solutions of the present application, the algorithms that can be used can be integrated, and this process is a process of performing fuzzy matching by using an integrated algorithm.

[0079] In the present implementation of the present application, after the fuzzy matching is performed by using the integration algorithm, some post filtering based on certain rules can be further performed, for example, the matching degree in the matching result is mapped to text description information, or the matching degree is properly compensated for or reduced based on a certain scenario.

[0080] Based on the previous description, more intuitively, an implementation of the present application further provides a schematic flowchart illustrating an implementation of performing fuzzy matching by using an integrated algorithm in a method for matching names in an actual application scenario, as shown in FIG. 3. In FIG. 3, a sequence of performing various detections and calculating matching degrees in various methods is merely an example, and is not for limiting the present application.

[0081] In FIG. 3, if matching of a name to be matched succeeds through any one

of the previous detections, a matching result can be directly output; otherwise, the matching result of the name to be matched can be determined and output by using one or more methods for matching degree calculation (for example, text matching degree calculation, phonetic matching degree calculation, and string matching degree calculation).

[0082] Certainly, various detections and matching degree calculation in various methods can also be completely performed, and then various detection results and matching degree calculation results in various methods can be comprehensively considered, to determine a matching result of a string to be matched.

[0083] Further, an implementation of the present application further provides a schematic flowchart illustrating an implementation of a method for matching names in an actual application scenario, as shown in FIG. 4. In FIG. 4, predetermining based on certain rules can include the following: determining whether a name to be matched is not included in a set of names that do not need to be matched, where a scan list index is an index of names in the standard name set.

[0084] Steps in FIG. 3 and FIG. 4 are described above in detail, and details are omitted here for simplicity.

[0085] The method for matching names provided in the present implementation of the present application is described above. As shown in FIG. 5, based on the same invention idea, an implementation of the present application further provides a corresponding apparatus.

[0086] FIG. 5 is a schematic structural diagram illustrating an apparatus for matching names and corresponding to FIG. 1, according to an implementation of the present application. A dashed line represents an optional module, and the apparatus includes the following: an acquisition module 501, configured to obtain a name to be matched; a determining module 502, configured to determine a standard name set used to match the name to be matched; a detection module 503, configured to detect the name to be matched to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical; and a matching module 504, configured to determine a matching result of the name to be matched based on a detection result.

[0087] Optionally, prior to determining the standard name set used to match the name to be matched, the determining module 502 is configured to obtain a predetermined set of names that do not need to be matched, and determine that the

name to be matched is not included in the set of names that do not need to be matched.

[0088] Optionally, that the determining module 502 is configured to determine a standard name set used to match the name to be matched includes the following:

5 **[0089]** The determining module 502 is configured to obtain a first name set that can be used to match the name to be matched, and determine the standard name set used to match the name to be matched by performing similarity matching on each word included in the name to be matched and each word included in a name in the first name set.

10 **[0090]** Optionally, that the determining module 502 is configured to determine the standard name set used to match the name to be matched by performing similarity matching on each word included in the name to be matched and each word included in a name in the first name set includes the following:

[0091] The determining module 502 is configured to obtain an index of each
15 name included in the first name set, where the index of the name is any word included in the name; segment the name to be matched to obtain each word included in the name to be matched; perform similarity matching on each word included in the name to be matched and each index to obtain a subset of the first name set. The obtained subset includes a name indexed by each index that is successfully matched; and
20 determine the standard name set used to match the name to be matched based on the subset.

[0092] Optionally, that the determining module 502 is configured to perform similarity matching on each word included in the name to be matched and each index includes the following:

25 **[0093]** The determining module 502 is configured to perform similarity matching on each word included in the name to be matched and each index by using a string matching algorithm, where the string matching algorithm includes at least one of the following: a prefix tree matching algorithm, a dictionary tree matching algorithm, a string similarity matching algorithm, or a pronunciation similarity matching
30 algorithm.

[0094] Optionally, the apparatus further includes the following: an alignment module 505, configured to align each word included in a name in the standard name set with each word included in the name to be matched based on the similarity between each word included in the name in the standard name set and each word

included in the name to be matched prior to the detection module 503 detects the name to be matched.

[0095] That the detection module 503 is configured to detect the name to be matched includes the following:

- 5 **[0096]** The detection module 503 is configured to detect the name to be matched based on the aligned standard name set to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

- 10 **[0097]** Optionally, that the detection module 503 is configured to detect the name to be matched includes the following: the detection module 503 is configured to perform at least one of abbreviation detection, address term detection, multi-language detection, or alias detection on the name to be matched to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

- 15 **[0098]** Optionally, that the detection module 503 is configured to perform abbreviation detection on the name to be matched includes the following:

- [0099]** The detection module is configured to obtain predetermined abbreviation contrast combination data, where each abbreviation contrast combination reflects an abbreviation mapping relationship between at least one word and an abbreviation of the word; detect whether a word included in the name to be matched has the abbreviation mapping relationship with a word included in a name in the standard name set based on the abbreviation contrast combination data; and determine, based on a detection result, whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

- 25 **[0100]** Optionally, that the detection module 503 is configured to perform address term detection on the name to be matched includes the following:

- [0101]** The detection module 503 is configured to obtain predetermined address term data; detect whether the name to be matched includes the address term, where it is considered that the address term does not affect a meaning of the name that includes the address term; and determine, based on a detection result, whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

[0102] Optionally, the alias includes a nickname of a name that corresponds to the alias, or a synonymous name of a name that corresponds to the alias in different fields.

That the detection module 503 is configured to perform alias detection on the name to be matched includes the following:

[0103] The detection module 503 is configured to perform at least one of the following: detecting a nickname of the name to be matched, or detecting a
5 synonymous name of the name to be matched in different fields.

[0104] Optionally, that the detection module 503 is configured to perform multi-language detection on the name to be matched includes the following: the detection module 503 is configured to determine a language that corresponds to the name to be matched; obtain at least one of a spelling deformation synonym rule or a
10 spelling deformation homonym rule of the language; and detect the name to be matched according to at least one of the spelling deformation synonym rule or the spelling deformation homonym rule, to determine whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

15 **[0105]** Optionally, that the detection module 503 is configured to perform alias detection on the name to be matched includes the following:

[0106] The detection module 503 is configured to obtain predetermined alias contrast combination data, where each alias contrast combination reflects an alias mapping relationship between at least one word and an alias of the word; detect
20 whether a word included in the name to be matched has the alias mapping relationship with a word included in a name in the standard name set based on the alias contrast combination data; and determine, based on a detection result, whether the name to be matched is synonymous with at least one name in the standard name set but characters of the names are not identical.

25 **[0107]** Optionally, that the matching module 504 is configured to determine a matching result of the name to be matched based on a detection result includes the following:

[0108] The matching module 504 is configured to determine the at least one name as the matching result of the name to be matched if the detection module 503
30 determines that the name to be matched is synonymous with the at least one name in the standard name set but the characters of the names are not identical; or otherwise, determine the matching result of the name to be matched by matching the name to be matched with a name in the standard name set by using one or more similarity algorithms.

[0109] Optionally, the similarity algorithm includes at least one of the following: an algorithm used to calculate a text matching degree, an algorithm used to calculate a phonetic matching degree, or an algorithm used to calculate a string matching degree.

[0110] Optionally, the name is a person name.

5 **[0111]** The apparatuses provided in the implementations of the present application are in a one-to-one correspondence with the methods. Therefore, the apparatuses and the methods have the similar beneficial technical effects. The beneficial technical effects of the methods have been described above in detail, and therefore the beneficial technical effects of the apparatuses are omitted here for simplicity.

10 **[0112]** In addition, application scenarios of the previous apparatuses and methods are not limited in the present application. In addition to the risk control field (for example, fields such as anti-money laundering and user authentication) mentioned in the background, the solutions of the present application are applicable to any other fields that may need to use the name matching technology.

15 **[0113]** In the 1990s, whether a technical improvement is a hardware improvement (for example, an improvement to a circuit structure, such as a diode, a transistor, or a switch) or a software improvement (an improvement to a method procedure) can be clearly distinguished. However, as technologies develop, current improvements to many method procedures can be considered as direct improvements to hardware
20 circuit structures. A designer usually programs an improved method procedure into a hardware circuit, to obtain a corresponding hardware circuit structure. Therefore, a method procedure can be improved by using a hardware entity module. For example, a programmable logic device (PLD) (for example, a field programmable gate array (FPGA)) is such an integrated circuit, and a logical function of the PLD is determined
25 by a user through device programming. The designer performs programming to "integrate" a digital system to a PLD without requesting a chip manufacturer to design and produce an application-specific integrated circuit chip. In addition, at present, instead of manually manufacturing an integrated circuit chip, such programming is mostly implemented by using "logic compiler" software. The logic compiler software
30 is similar to a software compiler used to develop and write a program. Original code needs to be written in a particular programming language for compilation. The language is referred to as a hardware description language (HDL). There are many HDLs, such as the Advanced Boolean Expression Language (ABEL), the Altera Hardware Description Language (AHDL), Confluence, the Cornell University

Programming Language (CUPL), HDCL, the Java Hardware Description Language (JHDL), Lava, Lola, MyHDL, PALASM, and the Ruby Hardware Description Language (RHDL). The very-high-speed integrated circuit hardware description language (VHDL) and Verilog are most commonly used. A person skilled in the art
 5 should also understand that a hardware circuit that implements a logical method procedure can be readily obtained once the method procedure is logically programmed by using the several described hardware description languages and is programmed into an integrated circuit.

[0114] A controller can be implemented by using any appropriate method. For
 10 example, the controller can be a microprocessor or a processor and a computer readable medium storing computer readable program code (such as software or firmware) that can be executed by the microprocessor or the processor, a logic gate, a switch, an application-specific integrated circuit (ASIC), a programmable logic controller, or a built-in microcontroller. Examples of the controller include but are not
 15 limited to the following microprocessors: ARC 625D, Atmel AT91SAM, Microchip PIC18F26K20, and Silicone Labs C8051F320. The memory controller can also be implemented as a part of the control logic of the memory. A person skilled in the art also knows that, in addition to implementing the controller by using the computer readable program code, logic programming can be performed on method steps to
 20 allow the controller to implement the same function in forms of the logic gate, the switch, the application-specific integrated circuit, the programmable logic controller, and the built-in microcontroller. Therefore, the controller can be considered as a hardware component, and an apparatus configured to implement various functions in the controller can also be considered as a structure in the hardware component. Or the
 25 apparatus configured to implement various functions can even be considered as both a software module implementing the method and a structure in the hardware component.

[0115] The system, apparatus, module, or unit illustrated in the previous implementations can be implemented by using a computer chip or an entity, or can be
 30 implemented by using a product having a certain function. A typical implementation device is a computer. The computer can be, for example, a personal computer, a laptop computer, a cellular phone, a camera phone, a smartphone, a personal digital assistant, a media player, a navigation device, an email device, a game console, a tablet computer, a wearable device, or a combination of any of these devices.

[0116] For ease of description, the apparatus above is described by dividing functions into various units. Certainly, when the present application is implemented, the functions of the units can be implemented in one or more pieces of software and/or hardware.

5 **[0117]** A person skilled in the art should understand that an implementation of the present disclosure can be provided as a method, a system, or a computer program product. Therefore, the present disclosure can use a form of hardware only implementations, software only implementations, or implementations with a combination of software and hardware. Moreover, the present disclosure can use a
10 form of a computer program product that is implemented on one or more computer-usable storage media (including but not limited to a disk memory, a CD-ROM, an optical memory, etc.) that include computer-usable program code.

[0118] The present disclosure is described with reference to the flowcharts and/or block diagrams of the method, the device (system), and the computer program product
15 based on the implementations of the present disclosure. It is worthwhile to note that computer program instructions can be used to implement each process and/or each block in the flowcharts and/or the block diagrams and a combination of a process and/or a block in the flowcharts and/or the block diagrams. These computer program instructions can be provided for a general-purpose computer, a dedicated computer, a
20 built-in processor, or a processor of another programmable data processing device to generate a machine, so that the instructions executed by the computer or the processor of the another programmable data processing device generate a device for implementing a specified function in one or more processes in the flowcharts and/or in one or more blocks in the block diagrams.

25 **[0119]** These computer program instructions can be stored in a computer readable memory that can instruct the computer or the another programmable data processing device to work in a specific way, so that the instructions stored in the computer readable memory generate an artifact that includes an instruction device. The instruction device implements a specified function in one or more processes in the
30 flowcharts and/or in one or more blocks in the block diagrams.

[0120] These computer program instructions can be loaded onto the computer or another programmable data processing device, so that a series of operations and operations and steps are performed on the computer or the another programmable device, thereby generating computer-implemented processing. Therefore, the

instructions executed on the computer or the another programmable device provide steps for implementing a specified function in one or more processes in the flowcharts and/or in one or more blocks in the block diagrams.

[0121] In a typical configuration, a computing device includes one or more processors (CPU), one or more input/output interfaces, one or more network
5 interfaces, and one or more memories.

[0122] The memory can include a non-persistent memory, a random access memory (RAM), a non-volatile memory, and/or another form that are in a computer readable medium, for example, a read-only memory (ROM) or a flash memory (flash
10 RAM). The memory is an example of the computer readable medium.

[0123] The computer readable medium includes persistent, non-persistent, movable, and unmovable media that can store information by using any method or technology. The information can be a computer readable instruction, a data structure, a program module, or other data. Examples of the computer storage medium include
15 but are not limited to a phase-change random access memory (PRAM), a static random access memory (SRAM), a dynamic random access memory (DRAM), another type of random access memory (RAM), a read-only memory (ROM), an electrically erasable programmable read-only memory (EEPROM), a flash memory or another memory technology, a compact disc read-only memory (CD-ROM), a digital
20 versatile disc (DVD) or another optical storage, a cassette magnetic tape, a magnetic tape/magnetic disk storage or another magnetic storage device, or any other non-transmission medium. The computer storage medium can be used to store information accessible by the computing device. Based on the description in the present specification, the computer readable medium does not include transitory
25 computer readable media (transitory media) such as a modulated data signal and carrier.

[0124] It is worthwhile to further note that, the terms "include", "comprise", or their any other variants are intended to cover a non-exclusive inclusion, so a process, a method, a product, or a device that includes a list of elements not only includes
30 those elements but also includes other elements which are not expressly listed, or further includes elements inherent to such process, method, product, or device. Without more constraints, an element preceded by "includes a ..." does not preclude the existence of additional identical elements in the process, method, product, or device that includes the element.

[0125] A person skilled in the art should understand that an implementation of the present application can be provided as a method, a system, or a computer program product. Therefore, the present application can use a form of hardware only implementations, software only implementations, or implementations with a combination of software and hardware. In addition, the present application can use a form of a computer program product that is implemented on one or more computer-usable storage media (including but not limited to a disk memory, a CD-ROM, an optical memory, etc.) that include computer-usable program code.

[0126] The present application can be described in the general context of computer executable instructions executed by a computer, for example, a program module. Generally, the program module includes a routine, a program, an object, a component, a data structure, etc. executing a particular task or implementing a particular abstract data type. The present application can also be practiced in distributed computing environments. In the distributed computing environments, tasks are performed by remote processing devices connected through a communications network. In a distributed computing environment, the program module can be located in both local and remote computer storage media including storage devices.

[0127] The implementations of the present specification are described in a progressive way. For the same or similar parts of the implementations, references can be made to the implementations. Each implementation focuses on a difference from other implementations. Particularly, a system implementation is basically similar to a method implementation, and therefore is described briefly. For related parts, references can be made to related descriptions in the method implementation.

[0128] The previous descriptions are implementations of the present application, and are not intended to limit the present application. A person skilled in the art can make various modifications and changes to the present application. Any modification, equivalent replacement, or improvement made without departing from the spirit and principle of the present application shall fall within the scope of the claims in the present application.

CLAIMS

1. A computer-implemented method for matching names, the computer-implemented method comprising:

receiving an input comprising a name to be matched, the name comprising a plurality of words;

determining a first name set corresponding to the name, the first name set comprising a plurality of elements;

performing word segmentation matching on the name to be matched, comprising:

obtaining an index of each name included in the first name set, wherein the index comprises a pre-established word associated with the name, and

segmenting the name to be matched to obtain each word included in the name to be matched;

performing similarity matching on each word included in the name to be matched and each index to obtain a respective subset of the first name set, by performing a plurality of types of string matching algorithms and, for each combination of string matching algorithm type and word included in the name to be matched, obtaining a respective subset, each subset comprising those names in the first name set that comprise a word that similarity matches the respective word of the name to be matched according to the respective type of string matching algorithm;

determining a standard name set from the subsets, comprising either:

(i) taking a union of the N subsets for each word in the name to be matched to determine M union sets, where M is a number words in the name to be matched and N is a number of types of string matching algorithms, and selecting each name whose total number of occurrences in the M union sets exceeds a specified threshold determining as the standard name set the names selected based on the total number of occurrences, or

(ii) obtaining N subsets, one for each type of string matching algorithm, intersecting the N subsets, and taking set as the standard name set, or wherein the standard name set includes only names similar with respect to at least one word and according to at least one string matching algorithm to the name to be matched; or

(iii) selecting as the names in the standard name set a portion of names that occur in at least two of the M union sets;

determining that the name is synonymous with at least one standard name in the standard name set despite that one or more name characters of the name are different from one or more standard name characters of the standard name, comprising performing at least one of abbreviation detection, address term detection, multi-language detection, or alias detection on the name to be matched to determine whether the name to be matched is synonymous with the at least one standard name in the standard name set; and

in response to determining that the name is synonymous with the at least one standard name, generating a matching result of the name comprising a weighted summation of a plurality of matching degrees of each of the plurality of words, the plurality of matching degrees comprising text matching degrees, phonetic matching degrees, and string matching degrees.

2. The computer-implemented method of claim 1, further comprising:

retrieving a predetermined set of names that do not need to be matched; and

determining that the name to be matched is not comprised in the predetermined set of names.

3. The computer-implemented method of claim 1, wherein taking a union of the N subsets for each word in the name to be matched to determine M union sets generates a comprehensive matching result of each word .

4. The computer-implemented method of claim 1, wherein the string matching algorithm comprises one or more of: a prefix tree matching algorithm, a dictionary tree matching algorithm, a string similarity matching algorithm, and a pronunciation similarity matching algorithm.

5. The computer-implemented method of claim 1, further comprising:

aligning, as an aligned standard name set, each word comprised in a name in the standard name set with each word comprised in the name to be matched based on a similarity between each word comprised in the standard name in the standard name set and each word comprised in the name to be matched; and

wherein determining whether the name is synonymous with the at least one standard name comprises:

detecting the name to be matched based on the aligned standard name set to determine

whether the name to be matched is synonymous with the at least one standard name in the standard name set.

6. The computer-implemented method of claim 1, wherein performing the abbreviation detection on the name to be matched comprises:

retrieving abbreviation contrast combination data, wherein each abbreviation contrast combination reflects an abbreviation mapping relationship between at least one word and an abbreviation of the word;

detecting whether a word comprised in the name to be matched has the abbreviation mapping relationship with a word comprised in the standard name in the standard name set based on the abbreviation contrast combination data; and

determining, based on a detection result, whether the name to be matched is synonymous with the at least one standard name in the standard name set.

7. The computer-implemented method of claim 1, wherein performing the address term detection on the name to be matched comprises:

retrieving address term data;

detecting whether the name to be matched comprises the address term, wherein it is considered that the address term does not affect a meaning of the name that comprises the address term; and

determining, based on a detection result, whether the name to be matched is synonymous with the at least one standard name in the standard name set.

8. The computer-implemented method of claim 1, wherein alias comprises a nickname of a name that corresponds to the alias or a synonymous name of a name that corresponds to the alias in different fields.

9. The computer-implemented method of claim 1, wherein performing the alias detection on the name to be matched comprises at least one of the following:

detecting a nickname of the name to be matched, or detecting a synonymous name of the name to be matched in different fields.

10. The computer-implemented method of claim 1, wherein performing the multi-language detection on the name to be matched comprises:

determining a language that corresponds to the name to be matched;

retrieving at least one of a spelling deformation synonym rule or a spelling deformation homonym rule of at least one of the language or another language; and

detecting the name to be matched according to at least one of the spelling deformation synonym rule or the spelling deformation homonym rule to determine whether the name to be matched is synonymous with the at least one standard name in the standard name set.

11. The computer-implemented method of claim 1, wherein performing the alias detection on the name to be matched comprises:

retrieving alias contrast combination data, wherein each alias contrast combination reflects an alias mapping relationship between at least one word and an alias of the word;

detecting whether a word comprised in the name to be matched has the alias mapping relationship with a word comprised in the standard name in the standard name set based on the alias contrast combination data; and

determining, based on a detection result, whether the name to be matched is synonymous with the at least one standard name in the standard name set.

12. The computer-implemented method of claim 11, wherein generating the matching result comprises:

in response to determining that the name to be matched is synonymous with the at least one standard name in the standard name set, determining the at least one standard name as the matching result of the name to be matched; or

in response to determining that the name to be matched is not synonymous with the at least one standard name in the standard name set, determining the matching result of the name to be matched by matching the name to be matched with the standard name in the standard name set by using one or more similarity algorithms.

13. The computer-implemented method of claim 12, wherein the similarity algorithm

comprises at least one of: an algorithm used to calculate a text matching degree, an algorithm used to calculate a phonetic matching degree, or an algorithm used to calculate a string matching degree.

14. An apparatus for matching names, the apparatus comprising a plurality of modules configured to perform the method of any one of claims 1 to 13.