

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.