

METHOD AND APPARATUS FOR OBTAINING ABBREVIATED NAME OF POINT OF INTEREST ON MAP

[0001] The present application claims priority to Chinese Patent Application No. 201610015852.7, filed on January 11, 2016 and entitled "METHOD AND
5 APPARATUS FOR OBTAINING ABBREVIATED NAME OF POINT OF
INTEREST ON MAP", which is incorporated here by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to computer technologies, and in particular,
to methods and apparatuses for obtaining an abbreviated name of a point of interest on
10 a map.

BACKGROUND

[0003] A point of interest (POI) on a map can be a house, a shop, a school, a bus
stop, etc. When the map is searched for a POI, a commonly used search method is to
search for the POI by using an abbreviated name of the POI. For example, when the
15 map is searched for "Beijing Normal University", an abbreviated name "BNU" of
"Beijing Normal University" can be alternatively entered to locate the university on
the map. Therefore, to support the method for searching a POI by using an
abbreviated name of the POI, some abbreviated names corresponding to a full name
of the POI need to be obtained to establish a mapping relationship between the full
20 name and the abbreviated name. As such, a corresponding full name can be
determined accurately based on an abbreviated name entered by a map user, and a
target address searched for by the user can be fed back to the user. In the existing
technology, an abbreviated name of a point of interest is usually obtained through
manual labeling, but the workload for manual labeling can be relatively high and
25 efficiency is relatively low. As the data volume of a map increases, the efficiency of
the manual labeling can become even lower.

SUMMARY

[0004] To resolve problems in the existing technology, the present disclosure provides a method and an apparatus for obtaining an abbreviated name of a point of interest on a map, so as to improve efficiency of obtaining an abbreviated name of a point of interest on a map.

[0005] According to a first aspect of implementations of the present disclosure, a method for obtaining an abbreviated name of a point of interest on a map is provided, including: generating a plurality of abbreviated names to be evaluated based on a full name of a point of interest on a map; obtaining, from a predetermined area of the point of interest, a plurality of address names including the full name of the point of interest or any of the abbreviated names to be evaluated; calculating a phrase status vector of a target phrase based on the address name, where the phrase status vector is used to indicate a location status of the target phrase in the address name, and the target phrase is the full name of the point of interest or any of the abbreviated names to be evaluated; separately calculating similarity between a phrase status vector of the full name of the point of interest and a phrase status vector of any of the abbreviated names to be evaluated; and determining an abbreviated name to be evaluated corresponding to similarity greater than a predetermined threshold as an abbreviated name of the point of interest corresponding to the full name of the point of interest.

[0006] According to a second aspect of the implementations of the present disclosure, an apparatus for obtaining an abbreviated name of a point of interest on a map is provided, including: a to-be-evaluated abbreviated name generation module, configured to generate a plurality of abbreviated names to be evaluated based on a full name of a point of interest on a map; a related address acquisition module, configured to obtain, from a predetermined area of the point of interest, a plurality of address names including the full name of the point of interest or any of the abbreviated names to be evaluated; a phrase status calculation module, configured to calculate a phrase status vector of a target phrase based on the address name, where the phrase status vector is used to indicate a location status of the target phrase in the address name, and the target phrase is the full name of the point of interest or any of the abbreviated names to be evaluated; a similarity calculation module, configured to separately calculate similarity between a phrase status vector of the full name of the point of interest and a phrase status vector of any of the abbreviated names to be evaluated;

and an abbreviated name determining module, configured to determine an abbreviated name to be evaluated corresponding to similarity greater than a predetermined threshold as an abbreviated name of the point of interest corresponding to the full name of the point of interest.

5 [0007] The technical solutions provided in the implementations of the present disclosure can include the following beneficial effects: By using the method, when a full name of a point of interest is provided, a related address name near the point of interest can be automatically obtained from a map address library, and an abbreviated name that can be an abbreviated name of the full name in all possible abbreviated
10 names to be evaluated is calculated based on the address name, so as to automatically determine the abbreviated name, thereby improving efficiency of obtaining an abbreviated name of a point of interest on a map.

[0008] It should be understood that the previous general description and the following detailed description are merely examples and explanations, and are not
15 intended to limit the present disclosure.

BRIEF DESCRIPTION OF DRAWINGS

[0009] The accompanying drawings here are incorporated in the specification, constitute a part of the specification, show implementations that are in accordance with the present disclosure, and are used with the specifications to illustrate principles
20 of the present disclosure.

[0010] FIG. 1 is a flowchart illustrating a method for obtaining an abbreviated name of a point of interest on a map, according to an example implementation;

[0011] FIG. 2 shows a principle of calculating cosine similarity in two-dimensional space, according to an example implementation; and

25 [0012] FIG. 3 shows an apparatus for obtaining an abbreviated name of a point of interest on a map, according to an example implementation.

DESCRIPTION OF IMPLEMENTATIONS

[0013] Example implementations are described here in detail, and the examples are presented in the accompanying drawings. When the following description refers to
30 the accompanying drawings, unless specified otherwise, same numbers in different

drawings represent the same or similar elements. Implementations described in the following example implementations do not represent all implementations consistent with the present disclosure. On the contrary, they are only examples of apparatuses and methods as described in the appended claims and that are consistent with some aspects of the present disclosure.

[0014] The implementations of the present application provide a method for obtaining an abbreviated name of a point of interest (POI) on a map. The method is a solution in which a corresponding abbreviated name is automatically obtained based on a full name of a POI without being manually labeled, so as to improve efficiency of obtaining the abbreviated name.

[0015] The abbreviated name acquisition method is based on the following principles: When a POI is used for naming an address, the full name and the abbreviated name of a POI appear in substantially similar locations in the detailed addresses. The full name and the abbreviated name are usually used in address names that correspond to the same location or geographically proximate locations.

[0016] For example, the POI is a school. Assuming that a full name of the school is Beijing Normal University, there are usually many address names using the school name inside or around the university, for example, Beijing Normal University Cafeteria, BNU Bookstore, Beijing Normal Dormitory Section A, BN Education Center, etc. (It should be noted that the previous names are only examples used for describing the solution of the present application and may not actually exist.) It can be seen that some address names use the full name, for example, "Beijing Normal University" in "Beijing Normal University Cafeteria"; and some address names use abbreviated names, for example, "BNU" in "BNU Bookstore", and "BN" in "BN Education Center".

[0017] In the previous example, a full name or an abbreviated name of a POI (POI) is applied to an address name, and the full name and the abbreviated name satisfy the two aspects mentioned above.

[0018] First, the full name and the abbreviated name are used in address names that correspond to the same location or geographically proximate locations. For example, "Beijing Normal University Cafeteria", "BNU Bookstore", and "Beijing Normal Dormitory Section A" are all located inside or around "Beijing Normal University".

[0019] Second, the full name and the abbreviated name of a POI appear in

substantially similar locations in the detailed addresses. "Beijing Normal University" is used as an example. For the abbreviated name "BNU" of "Beijing Normal University", if a detailed address string is described as "19 Xijiekou Street, Beijing Normal University Library, Beijing", in an address library, there may be an address string described as "19 Xijiekou Street, BNU Library, Beijing", or a string similar to "Xijiekou Street, BNU Library, Beijing". It shows that locations of the full name and the abbreviated name of the POI are substantially similar in the address library, and in other words, the context is substantially similar.

[0020] For example, for a large-scale address text library, a pair of full name and abbreviated name can have same phrases before and after. For example, in the address library of the previous example, the full name "Beijing Normal University" and the corresponding abbreviated name "BNU" all have phrases "Xijiekou Street" before and phrase "Library" after. For example, phrases before and after a pair of full name and abbreviated name include "Cafeteria", "Library", and "Building". These phrases can form a vector {Cafeteria, Library, Building}, and each phrase is a dimension of the vector.

Table 1 Vector dimensions corresponding to a full name and an abbreviated name

Dimension	Full name: Beijing Normal University	Abbreviated name: BNU	
Cafeteria	117	32	
Library	267	71	
Building	422	112	
	Cosine similarity		0.999977577

[0021] In the example shown in Table 1, "Cafeteria" appears 117 times in detailed address strings including the full name, and "Cafeteria" appears 32 times in detailed address strings including the abbreviated name. Similarly, for the full name and the abbreviated name, "Library" respectively appears 267 times and 71 times, and "Building" respectively appears 422 times and 112 times. It can be seen that numbers of appearances of the three dimensions are basically proportional. That is, a vector

corresponding to the full name "Beijing Normal University" is {117, 267, 422}, and a vector corresponding to the abbreviated name "BNU" is {32, 71, 112}. If cosine similarity between the two vectors is calculated, it can be seen that the cosine value is very close to 1.

5 **[0022]** Based on the previous principle, for a procedure of the method for obtaining an abbreviated name of a POI, refer to FIG. 1. The method includes the steps below.

[0023] Step 101: Generate a plurality of abbreviated names to be evaluated based on a full name of a POI on a map.

10 **[0024]** In the step, single words included in the full name of the POI can be combined in a single word arrangement sequence, to form combinations of any number of single words, where each combination is an abbreviated name, and a plurality of abbreviated names are obtained.

[0025] "Beijing Normal University" is still used as an example. The full name in Chinese Pinyin includes six single words: "Bei", "Jing", "Shi", "Fan", "Da", and "Xue". Then these single words are combined to form combinations of any number of single words. When a number of single words is 2, two single words are combined, for example, "Beijing", "Jingshi", "Shifan", "Fanda", "Daxue", "Beishi", "Beifan", etc. It can be seen that during the combination, the single words are combined in arrange
20 sequences of the single words in the full name. That is, the sequence of each single word in the combination is the same as the sequence of the single word in the full name. When the number of single words is 3, three single words are combined, for example, "Beijingshi", "Jingshifan", "Beishifan", etc.

[0026] After a plurality of abbreviated names are obtained through arbitrary
25 combination, a single word and a noise phrase determined as an abbreviated name of a non-POI can be further removed from the plurality of abbreviated names, and a remaining abbreviated name is the to-be-evaluated abbreviated name. As such, a quantity of abbreviated names to be evaluated participating in subsequent processing steps is decreased, thereby improving processing efficiency of the procedure. Single
30 words are the previous "Bei", "Jing", "Shi", etc., and the noise phrase includes "Beijing", "Shifan", "Daxue", etc. It is usually impossible for a user to enter these commonly used phrases as abbreviated names of "Beijing Normal University".

[0027] Step 102: Obtain, from a predetermined area of the POI, a plurality of address names including the full name of the POI or any of the abbreviated names to

be evaluated.

[0028] For example, the predetermined area can be an area within 500 meters near the POI, such as "Beijing Normal University Cafeteria", "BNU Bookstore", "Beijing Normal Dormitory Section A", and "BN Education Center" enumerated above. These address names include either the full name "Beijing Normal University" of the POI or an abbreviated name to be evaluated such as "BNU". The abbreviated names to be evaluated are the abbreviated names of the POI on the map determined in step 101.

[0029] Step 103: Calculate a phrase status vector of a target phrase based on the address name, where the phrase status vector is used to indicate a location status of the target phrase in the address name, and the target phrase is the full name of the POI or any of the abbreviated names to be evaluated.

[0030] The phrase status vector calculated in the step can be a multidimensional vector, for example, an n-dimensional vector $(A_1, A_2, \dots, A_n)$. In addition, a corresponding phrase status vector can be obtained through calculation for each of the full name of the POI and any of the abbreviated names to be evaluated. For example, a phrase status vector corresponding to the full name "Beijing Normal University" is $(A_1, A_2, \dots, A_n)$, and a phrase status vector corresponding to one of the abbreviated names "BNU" is $(B_1, B_2, \dots, B_n)$.

[0031] The phrase status vector can be calculated with reference to the plurality of address names obtained in step 102, and a method for calculating the phrase status vector of the full name of the POI is the same as that for calculating the phrase status vector of the abbreviated name to be evaluated. If the full name of the POI or any of the abbreviated names is referred to as the target phrase, the phrase status vector of one of the target phrases can be calculated in the following method:

[0032] A plurality of address names can be obtained in step 102, for example, "BNU Bookstore", "BN Education Center", etc. For a target phrase, for example, the target phrase is "BNU", address names including "BNU" are found, such as "BNU Bookstore". There can be a plurality of address names including the target phrase. A neighboring phrase of the target phrase can be obtained from each address name including the target phrase, and the neighboring phrases include phrases before and after.

[0033] For example, in the address name "BNU Bookstore" including the target phrase "BNU", the phrase before the target phrase is "null" (that is, there is no phrase previous to the target phrase), and the phrase after is "Bookstore". As another example,

in an address name "Bowen BNU Bookstore" including the target phrase "BNU", the phrase before the target phrase is "Bowen", and the phrase after is "Bookstore". After statistics collection, in all address names including the target phrase, there can be a plurality of neighboring phrases of the target phrase, such as the previously
5 enumerated "null", "Bookstore", "Bowen", and "Bookstore". Assume that only the four neighboring phrases (there can be more neighboring phrases in practice), the four phrases each can correspond to a vector dimension of the phrase status vector of the target phrase, and the vector (null, Bookstore, Bowen, Bookstore) is a four-dimensional vector.

10 **[0034]** A dimension value of phrase can be the frequency that the phrase appears in the address names. Using "Bookstore" as an example. Assume that 10 address names are obtained in step 102, and "Bookstore" appears twice, the frequency or dimension value of "Bookstore" is 2. Other dimension values can be similarly calculated and is not described again.

15 **[0035]** Step 104: Separately calculate similarity between a phrase status vector of the full name of the POI and a phrase status vector of any of the abbreviated names to be evaluated.

[0036] For example, after the phrase status vector of the full name of the POI and the phrase status vector of any of the abbreviated names to be evaluated are calculated
20 in step 103, similarity between the phrase status vector of the full name of the POI and the phrase status vector of each abbreviated name to be evaluated is calculated. The similarity calculated in the example can be cosine similarity.

[0037] Referring to FIG. 2, FIG. 2 shows calculation of cosine similarity in two-dimensional space, and a and b are two two-dimensional vectors. It is assumed
25 that coordinates of a are (x1, y1) and coordinates of b are (x2, y2). The two vectors are drawn into vector space based on the coordinate values, as shown in FIG. 2. Then a cosine value corresponding to an angle between the two two-dimensional vectors a and b is calculated, and the cosine value can be used to represent similarity between the two vectors. If the two vectors are more similar, an angle between the two vectors
30 is smaller, the cosine value is closer to 1, and their directions are more similar.

[0038] The cosine value of the angle is calculated based on a vector dot product

equation expressed as follows:

$$\cos \theta = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \times \sqrt{x_2^2 + y_2^2}}$$

[0039] The previous equation can be extended to calculate cosine similarity between multidimensional vectors. If the phrase status vector of the full name of the POI is $A = (A_1, A_2, \dots, A_n)$, and the phrase status vector of one of the abbreviated names to be evaluated is $(B_1, B_2, \dots, B_n)$, cosine similarity between the two vectors is

$$\cos \theta = \frac{\sum_1^n (A_i + B_i)}{\sqrt{\sum_1^n A_i^2} \times \sqrt{\sum_1^n B_i^2}}$$

5 calculated as:

[0040] Step 105: Determine an abbreviated name to be evaluated corresponding to similarity greater than a predetermined threshold as an abbreviated name of the POI corresponding to the full name of the POI.

[0041] For example, if the predetermined threshold is 0.42, when a calculated value of the cosine similarity is greater than or equal to 0.42, an abbreviated name to be evaluated corresponding to the similarity can be determined as the abbreviated name of the POI corresponding to the full name of the POI. For example, if cosine similarity between a phrase status vector A of the full name "Beijing Normal University" of the POI and a phrase status vector B of one of the abbreviated names
10 "BNU" is 0.7, it can be determined that the abbreviated name "BNU" is used as the abbreviated name of the full name "Beijing Normal University".

[0042] In the example, there is at least one abbreviated name greater than the predetermined threshold. For example, the abbreviated name of the full name "Beijing Normal University" can include a plurality of abbreviated names, such as "BNU",
15 "BN", and "Beijing Normal".

[0043] In the method for obtaining an abbreviated name of a POI on a map, when a full name of a POI is provided, a related address name near the POI can be automatically obtained from a map address library, and an abbreviated name in all possible abbreviated names can be calculated based on the address, so as to
20 automatically determine the abbreviated name. As such, the efficiency of obtaining the abbreviated name can be improved. In addition, the method can be performed on a distributed computing platform, so as to further improve a processing speed.

[0044] After a mapping relationship between a full name and an abbreviated name of a POI is determined, the mapping relationship can be used to search a POI on a map, or another scenario related to address entry. For example, when a user enters
25 "Dongwai Avenue", it can be seen through comparison of a full name and an abbreviated name that a full name of "Dongwai Avenue" is "Dongzhimen Avenue".

Then the user can be asked whether to enter Dongzhimen Avenue. In addition, subsequent possible detailed addresses related to Dongzhimen Avenue are provided as auto-complete options for the user to select, for example, addresses such as restaurants and hotels near Dongzhimen Avenue.

5 **[0045]** The implementations of the present application further provide an apparatus for obtaining an abbreviated name of a POI on a map. As shown in FIG. 3, the apparatus can include a to-be-evaluated abbreviated name generation module 31, a related address acquisition module 32, a phrase status calculation module 33, a similarity calculation module 34, and an abbreviated name determining module 35.

10 **[0046]** The to-be-evaluated abbreviated name generation module 31 is configured to generate a plurality of abbreviated names to be evaluated based on a full name of a POI on a map.

[0047] The related address acquisition module 32 is configured to obtain, from a predetermined area of the POI, a plurality of address names including the full name of the POI or any of the abbreviated names to be evaluated.

15 **[0048]** The phrase status calculation module 33 is configured to calculate a phrase status vector of a target phrase based on the address name, where the phrase status vector is used to indicate a location status of the target phrase in the address name, and the target phrase is the full name of the POI or any of the abbreviated names to be evaluated.

20 **[0049]** The similarity calculation module 34 is configured to separately calculate similarity between a phrase status vector of the full name of the POI and a phrase status vector of any of the abbreviated names to be evaluated.

[0050] The abbreviated name determining module 35 is configured to determine an abbreviated name corresponding to similarity greater than a predetermined threshold as an abbreviated name of the POI corresponding to the full name of the POI.

25 **[0051]** In an example, when generating the plurality of abbreviated names to be evaluated based on the full name of the POI on the map, the to-be-evaluated abbreviated name generation module 31 is configured to: combine single words included in the full name of the POI in a single word arrangement sequence, to form combinations of any number of single words, where each combination is an abbreviated name, and a plurality of abbreviated names are obtained; and remove, from the plurality of abbreviated names, a single word and a noise phrase determined

as an abbreviated name of a non-POI, where a remaining abbreviated name is the to-be-evaluated abbreviated name.

5 [0052] In an example, when calculating the phrase status vector of the target phrase based on the address name, the phrase status calculation module 33 is configured to obtain a neighboring phrase of the target phrase from any of the plurality of address names, and use a frequency that each neighboring phrase appears in the plurality of address names as a dimension value of a vector dimension of the phrase status vector of the target phrase.

10 [0053] In an example, the similarity calculation module 34 is configured to calculate cosine similarity between the phrase status vector of the full name of the POI and the phrase status vector of any of the abbreviated names to be evaluated.

[0054] In an example, there is at least one to-be-evaluated abbreviated name corresponding to the similarity greater than the predetermined threshold.

15 [0055] A person skilled in the art can easily figure out another implementation solution of the present disclosure after considering the specifications and practicing the present disclosure. The present application is intended to cover any variations, functions, or adaptive changes of the present disclosure. These variations, functions, or adaptive changes comply with general principles of the present disclosure, and include common knowledge or a commonly used technical means in the technical
20 field that is not disclosed in the present application. The specification and the implementations are merely considered as examples, and the actual scope and the spirit of the present disclosure are described by the following claims.

[0056] It should be understood that the present disclosure is not limited to the structures that are described above and that are shown in the accompanying drawings,
25 and modifications and changes can be made without departing from the scope of the present disclosure. The protection scope of the present disclosure is defined by the appended claims.

CLAIMS

1. A computer implemented method comprising:

obtaining an abbreviated name of a point of interest on an electronic map, comprising:

generating a plurality of abbreviated names to be evaluated based on a full name of the point of interest on the electronic map;

obtaining, from an area within a predetermined distance of the point of interest, a plurality of address names comprising the full name of the point of interest or any of the abbreviated names to be evaluated;

calculating a phrase status vector of a target phrase based on the address name, wherein the phrase status vector is used to indicate a location status of the target phrase in the address name, and the target phrase is the full name of the point of interest or any of the abbreviated names to be evaluated;

separately calculating similarity between a phrase status vector of the full name of the point of interest and a phrase status vector of any of the abbreviated names to be evaluated; and

determining an abbreviated name to be evaluated corresponding to similarity greater than a predetermined threshold as an abbreviated name of the point of interest corresponding to the full name of the point of interest;

determining a mapping relationship between the full name of the point of interest and the determined abbreviated name of the point of interest;

receiving, from an electronic map user, an address search query comprising an abbreviated name of the point of interest;

identifying the full name of the point of interest using the established mapping relationship; and

providing the electronic map user with an address corresponding to the identified full name of the point of interest.

2. The method of claim 1, wherein generating a plurality of abbreviated names to be evaluated based on a full name of the point of interest on the electronic map comprises:

combining single words comprised in the full name of the point of interest in a single word arrangement sequence, to form combinations of any number of single words, wherein each combination is an abbreviated name, and a plurality of abbreviated names are obtained;

determining one or more single words and noise phrases that are not abbreviated names of the point of interest, comprising identifying one or more single words or noise phrases that an

electronic map user would not enter when searching for the point of interest; and

removing, from the plurality of abbreviated names, the determined single words and noise phrases, wherein a remaining abbreviated name is the to-be-evaluated abbreviated name.

3. The method of claim 1, wherein calculating a phrase status vector of a target phrase based on the address name, wherein the phrase status vector is used to indicate a location status of the target phrase in the address name comprises:

obtaining a neighboring phrase of the target phrase from any of the plurality of address names; and

using a frequency that each neighboring phrase appears in the plurality of address names as a dimension value of a vector dimension of the phrase status vector of the target phrase.

4. The method of claim 1, wherein calculating similarity between a phrase status vector of the full name of the point of interest and a phrase status vector of any of the abbreviated names to be evaluated comprises:

calculating cosine similarity between the phrase status vector of the full name of the point of interest and the phrase status vector of any of the abbreviated names to be evaluated.

5. The method of claim 1, wherein there is at least one to-be-evaluated abbreviated name corresponding to the similarity greater than the predetermined threshold.

6. The method of claim 1, further comprising;

providing, to the electronic map user, a prompt asking whether the electronic map user wants to enter the identified full name of the point of interest as a subsequent address search query.

7. The method of claim 6, further comprising providing, to the electronic map user, one or more user selectable auto-complete options comprising addresses related to the determined corresponding full name of the point of interest.

8. An apparatus comprising multiple modules configured to perform the method of any one of claims 1 to 7.