

DESCRIPTION

TITLE OF INVENTION: METHOD AND DEVICE FOR DE-IDENTIFYING LOCATION INFORMATION

Technical Field

5 The following description relates to a method and device for de-identifying location information, and more particularly, to technology for maximizing the anonymity and usability of de-identified data in de-identifying location information of an individual.

Background Art

 Recently, the number of services provided using the result of big data analytics
10 has been increasing. Various types of services using big data may provide convenience to users and the types may be infinite.

 However, since information such as address information, medical records, and private information of the users used as big data is personal information, the use of the information may be subject to legal restrictions, such as individual consent. Therefore,
15 technology for de-identifying personal information is required.

 In technology for de-identifying personal information, it may be possible that the usability of de-identified data decreases when the degree of de-identification is excessively high and that personal information is identified when the degree of de-identification is excessively low. Thus, such usability and anonymity should be
20 appropriately considered.

Disclosure of the Invention

Technical Goals

 The present disclosure may, in de-identifying location information such as an address of a user among pieces of personal information of users, provide a method and
25 device for de-identification that may maximize usability and anonymity by dividing an entire area into a plurality of unit areas and grouping adjacent unit areas according to an anonymity threshold value.

 In addition, the present disclosure may provide a method and device for de-identification that may allow a multi-dimensional analysis through generating a multi-
30 dimensional de-identified area by de-identifying location information to which different types of personal information are mapped.

 Furthermore, the present disclosure may provide a method and device for de-

identification that may control usability and anonymity by setting the unit area or anonymity threshold value differently depending on embodiments.

Technical Solutions

5 A method of de-identifying location information according to an embodiment of the present disclosure may include: identifying an anonymity threshold value and an entire area including pieces of location information of users; dividing the entire area into a plurality of unit areas; determining a number of pieces of location information included in each of the unit areas; and generating de-identified areas by grouping the unit areas based on the anonymity threshold value and the number of pieces of location information
10 included in each of the unit areas.

A number of the generated de-identified areas may increase as the anonymity threshold value decreases and decrease as the anonymity threshold increases.

The generating of the de-identified areas may, when a number of pieces of location information included in an arbitrary unit area among the unit areas is less than
15 the anonymity threshold value, determine the de-identified areas by grouping the arbitrary unit area with a unit area adjacent to the arbitrary unit area.

The generating of the de-identified areas may, when a number of pieces of location information included in a de-identified area is less than the anonymity threshold value, group the de-identified area with a unit area adjacent to the de-identified area.

20 A number of pieces of location information included in each of the generated de-identified areas is greater than or equal to the anonymity threshold value.

The generating of the de-identified areas may further include generating a plurality of combinations of de-identified areas by grouping the unit areas in a plurality of different ways and determining a combination of de-identified areas with a highest
25 usability evaluation result among the generated combinations of de-identified areas as a final de-identification result.

The generating of the de-identified areas may include generating a plurality of combinations of de-identified areas by grouping the unit areas in a plurality of different ways and determining a combination of de-identified areas with a number of the de-
30 identified areas less than or equal to a preset standard number among the generated combinations of de-identified areas as a final de-identification result.

A method of de-identifying location information according to an embodiment of

the present disclosure may include: identifying a geographic area including pieces of location information of users belonging to an arbitrary category; dividing the geographic area into a plurality of unit areas for each dimension; comparing a number of pieces of location information included in each of the unit areas to an anonymity threshold value that is preset; generating de-identified areas by grouping the unit areas based on the comparison result; and obtaining a multi-dimensional de-identified area by combining i) a de-identified area with ii) a de-identified area generated from a geographic area including pieces of location information of users belonging to a category other than the arbitrary category.

A number of the generated de-identified areas may increase as the anonymity threshold value decreases and decrease as the anonymity threshold increases.

The generating of the de-identified areas may, when a number of pieces of location information included in an arbitrary unit area among the unit areas is less than the anonymity threshold value, determine the de-identified areas by grouping the arbitrary unit area with a unit area adjacent to the arbitrary unit area.

The generating of the de-identified areas may, when a number of pieces of location information included in a de-identified area is less than the anonymity threshold value, group the de-identified area with a unit area adjacent to the de-identified area.

A number of pieces of location information included in each of the generated de-identified areas is greater than or equal to the anonymity threshold value.

A device for de-identifying location information according to an embodiment of the present disclosure may include a processor, wherein the processor may be configured to identify an anonymity threshold value and an entire area including pieces of location information of users, divide the entire area into a plurality of unit areas, determine a number of pieces of location information included in each of the unit areas, and generate de-identified areas by grouping the unit areas based on the anonymity threshold value and the number of pieces of location information included in each of the unit areas.

A number of the generated de-identified areas may increase as the anonymity threshold value decreases and decrease as the anonymity threshold increases.

When a number of pieces of location information included in an arbitrary unit area among the unit areas is less than the anonymity threshold value, the processor may determine the de-identified areas by grouping the arbitrary unit area with a unit area

adjacent to the arbitrary unit area.

When a number of pieces of location information included in a de-identified area is less than the anonymity threshold value, the processor may group the de-identified area with a unit area adjacent to the de-identified area.

- 5 A number of pieces of location information included in each of the generated de-identified areas is greater than or equal to the anonymity threshold value.

10 The processor may generate a plurality of combinations of de-identified areas by grouping the unit areas in a plurality of different ways and determine a combination of de-identified areas with a highest usability evaluation result among the generated combinations of de-identified areas as a final de-identification result.

15 The processor may generate a plurality of combinations of de-identified areas by grouping the unit areas in a plurality of different ways and determine a combination of de-identified areas with a number of the de-identified areas less than or equal to a preset standard number among the generated combinations of de-identified areas as a final de-identification result.

20 A device for de-identifying location information according to an embodiment of the present disclosure may include a processor, wherein the processor may be configured to identify a geographic area including pieces of location information of users belonging to an arbitrary category, divide the geographic area into a plurality of unit areas for each dimension, compare a number of pieces of location information included in each of the unit areas to an anonymity threshold value that is preset, generate de-identified areas by grouping the unit areas based on the comparison result, and obtain a multi-dimensional de-identified area by combining i) a de-identified area with ii) a de-identified area generated from a geographic area including pieces of location information of users belonging to a category other than the arbitrary category.

Effects

30 According to an embodiment of the present disclosure, in de-identifying location information such as an address of a user among pieces of personal information of users, usability and anonymity may be maximized by dividing an entire area into a plurality of unit areas and grouping adjacent unit areas according to an anonymity threshold value.

In addition, according to an embodiment of the present disclosure, a multi-dimensional analysis of location information may be possible through generating a multi-

dimensional de-identified area by de-identifying location information to which different types of personal information are mapped.

Furthermore, according to an embodiment of the present disclosure, usability and anonymity may be controlled to perform de-identification of personal information by
5 setting the unit area or anonymity threshold value differently depending on embodiments. In addition, according to an embodiment of the present disclosure, de-identified location information of users may be used when analyzing the time required for each delivery distance.

Brief Description of Drawings

10 FIG. 1 is a diagram illustrating a de-identification device according to an embodiment.

FIG. 2 is a diagram illustrating an example of an entire area and location information that is a de-identification target according to an embodiment.

15 FIG. 3 is a diagram illustrating an example of unit areas according to an embodiment.

FIG. 4 is a diagram illustrating a process of grouping unit areas, according to an embodiment.

FIG. 5 is a diagram illustrating a result of de-identifying location information of users belonging to different categories, according to an embodiment.

20 FIG. 6 is a flowchart illustrating a de-identification method according to an embodiment.

Best Mode for Carrying Out the Invention

Hereinafter, embodiments are described in detail with reference to the accompanying drawings. However, various alterations and modifications may be made
25 to the embodiments. Here, the embodiments are not meant to be limited by the descriptions of the present disclosure. The embodiments should be understood to include all changes, equivalents, and replacements within the idea and the technical scope of the disclosure.

The terminology used herein is for the purpose of describing particular
30 embodiments only and is not intended to be limiting. The singular forms "a," "an," and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It should be further understood that the terms "comprises/comprising" and/or

"includes/including," when used herein, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components and/or groups thereof.

5 Unless otherwise defined, all terms including technical or scientific terms used herein have the same meaning as those commonly understood by one of ordinary skill in the art to which the embodiments belong. Terms, such as those defined in commonly used dictionaries, are to be interpreted as having a meaning that is consistent with their meaning in the context of the relevant art and the present disclosure, and are not to be
10 interpreted in an idealized or overly formal sense unless expressly so defined herein.

When describing the embodiments with reference to the accompanying drawings, like reference numerals refer to like components and a repeated description related thereto is omitted. In the description of embodiments, detailed description of well-known related structures or functions is omitted when it is deemed that such description may
15 cause ambiguous interpretation of the present disclosure.

FIG. 1 is a diagram illustrating a de-identification device according to an embodiment.

The present disclosure relates to technology that, in de-identifying location information 105 such as an address of a user among personal information of the user, may
20 maximize usability and anonymity of the location information 105 that is de-identified by dividing an entire area 103 into a plurality of unit areas and grouping adjacent unit areas according to an anonymity threshold value.

Referring to FIG. 1, a de-identification device 101 may include a processor 102. The processor 102 of the de-identification device 101 may perform de-identification
25 methods according to various embodiments of the present disclosure. Referring to FIG. 1, the processor 102 may generate de-identified areas by de-identifying the entire area 103 including the location information 105 of individuals.

The location information 105 may refer to information on a location of a user on a map. For example, the location information 105 may include an address of the user
30 and may include the latitude, longitude, and height of the location of the user in three-dimensional space. The location information 105 may have different meanings depending on a type of service or an embodiment.

For example, the location information 105 of a user collected in connection with delivery and shipping services may include an address. The location information 105 of a user collected in an application used by a rider performing a delivery may include the latitude and longitude of a parking location related to a delivery address.

5 Accordingly, the location information 105 may not be limited to a specific example and may refer to the location of a user on a two-dimensional map or in three-dimensional space.

A user may refer to people who use an application or services provided on the web. The location information 105 of users belonging to an arbitrary category may be

10 de-identified using a de-identification method according to an embodiment.

For example, the users belonging to an arbitrary category may refer to users belonging to a category in which some aspects of personal information are common, such as men in their 20s and women in their 30s. When the location information 105 for multiple categories is de-identified, a multi-dimensional de-identified area in which de-

15 identified areas for each category are combined may be generated.

The entire area 103 may refer to a determined range of a two-dimensional or three-dimensional geographic area. The range of the entire area 103 is determined based on the range of the location information 105 that is a de-identification target, and the range of the entire area 103 may not be limited to a specific example. A specific

20 example of the entire area 103 is described below with reference to FIG. 2.

Referring to FIG. 1, the entire area 103 may include the location information 105 of multiple users. The de-identification device 101 may generate a de-identified area by de-identifying the location information 105 of the users. The de-identified area may refer to a determined range of area in an entire area 104 that the location information 105

25 of the users has, such as Area 1 and Area 2, in the entire area 104 of FIG. 1.

Each de-identified area may include the location information 105 of multiple users, but the location information 105 of each user may not be specified. According to an embodiment, as a number of pieces of the location information 105 included in the de-identified area increases, it may become difficult for each user to be specified and thus,

30 anonymity may increase.

According to an embodiment, the de-identified area may include the number of pieces of location information 105 that is equal to or greater than an anonymity threshold

value that is preset. The anonymity threshold value may refer to a value indicating a degree of de-identification. For example, the anonymity threshold value may be set to a positive natural number.

That is, as the anonymity threshold value increases, the anonymity of the location information 105 to be de-identified may increase, and as the degree of de-identification decreases, the anonymity of the location information 105 to be de-identified may decrease. Therefore, the anonymity threshold value may be set differently depending on embodiments or purpose of use and is not limited to a specific example.

For example, when the anonymity threshold value is "4," a de-identified area may be generated so that the number of pieces of location information 105 included in the de-identified area may be "4." A detailed process in which the de-identification device 101 generates a de-identified area is described below with reference to FIG. 3.

For example, the location information 105 that is de-identified may be used for big data analytics. For example, when providing a delivery service, by analyzing delivery times and data obtained by de-identifying addresses of users who used the delivery service, the time required according to a delivery distance may be predicted when a new delivery task occurs.

That is, in the example above, when the addresses of the users are de-identified by a village unit, it may be difficult to predict a specific time required, and when an address of a user is de-identified as is, the location may be specified and accordingly the user may be identified. Consequently, an issue related to personal information may arise. Therefore, the de-identification method of the present disclosure, which may maximize usability and anonymity, may be used.

FIG. 2 is a diagram illustrating an example of an entire area and location information that is a de-identification target according to an embodiment.

FIG. 2 is a diagram illustrating an example of an entire area 200 including pieces of location information 201. The entire area 200 may refer to a two-dimensional or three-dimensional geographic area. In the case of FIG. 2, the entire area 200 may be represented as a two-dimensional map.

For example, in FIG. 2, the location information 201 may include addresses of users and may include the latitude, longitude, and height of the location of a user in three-dimensional space. The location information 201 may have different meanings

depending on the type of service or an embodiment. The location information 201 may be de-identified according to various embodiments of the present disclosure because personal information of a user, such as an address, may be exposed.

A de-identification device may identify the entire area 200 including the location
5 information 201 of users and divide the entire area 200 into a plurality of unit areas. An example of divided unit areas is described below with reference to FIG. 3.

FIG. 3 is a diagram illustrating an example of unit areas according to an embodiment.

For example, FIGS. 3A to 3C may be diagrams illustrating entire areas 301 to
10 303 divided into unit areas 311 to 313 of different shapes on a two-dimensional map 300, according to various embodiments. According to various embodiments, a de-identification device may divide the entire areas 301 to 303 including pieces of location information into unit areas of various shapes.

FIG. 3A may illustrate a case in which the entire area 301 is divided into square-
15 shaped unit areas 311. FIG. 3B may illustrate a case in which the entire area 302 is divided into regular hexagonal-shaped unit areas 312. FIG. 3C may illustrate a case in which the entire area 303 is divided into polygonal-shaped unit areas 313.

For example, the de-identification device may separate an entire range of the entire areas 301 to 303 into the unit areas 311 to 313. That is, the de-identification
20 device may divide the entire areas 301 to 303 into the unit areas 311 to 313 without any gaps. However, the width or volume of all unit areas 311 to 313 may not be the same.

The de-identification device may divide the entire areas 301 to 303 into the unit areas 311 to 313 so that at least one edge or face may be shared between the unit areas 311 to 313 that are adjacent to each other. Depending on an embodiment, the unit areas
25 311 to 313 of various shapes may be used under the condition that at least one edge or face is shared between the unit areas 311 to 313 that are adjacent to each other. In another example, when an entire area in three-dimensional space is identified, the de-identification device may divide the entire area into unit areas of a three-dimensional polyhedron.

Each of the unit areas 311 to 313 may include pieces of location information of
30 multiple users. The pieces of location information included in each of the unit areas 311 to 313 may be de-identified as a de-identified area. The de-identification device may

generate de-identified areas by grouping the unit areas 311 to 313 based on the anonymity threshold value and the number of pieces of location information included in each of the unit areas 311 to 313. A detailed process of de-identification of the location information is described below with reference to FIG. 4.

5 FIG. 4 is a diagram illustrating a process of grouping unit areas, according to an embodiment.

For example, referring to FIG. 4, a de-identification device may identify an entire area 410 including pieces of location information of users on a two-dimensional map 400. For example, a processor of the de-identification device may load map data corresponding
10 to the entire area 410 and data on the pieces of location information of the users from a memory of the de-identification device.

The de-identification device may divide the entire area 410 into a plurality of unit areas 411. For example, referring to FIG. 4, the de-identification device may divide the entire area 410 into square-shaped unit areas 411.

15 The de-identification device may determine the number of pieces of location information included in each unit area 411. For example, the de-identification device may determine the number of pieces of location information that belong to each unit area 411 by considering the boundary of each unit area 411 and the latitude and longitude included in the location information.

20 For example, the number of pieces of location information included in the unit area 411 to which a residential area in which users may cluster such as an apartment or residential neighborhood belongs may be greater than the number of pieces of location information included in the unit area 411 to which a natural area such as a river or mountain belongs.

25 Referring to FIG. 4, the number of pieces of location information included in each unit area 411 of an entire area 420 may be determined. The de-identification device may identify an anonymity threshold value 401 that is preset. The de-identification device may generate de-identified areas 431 to 441 by grouping the unit areas 411 based on the anonymity threshold value 401 and the number of pieces of location information
30 included in each of the unit areas 411.

When the number of pieces of location information included in an arbitrary unit area 411 among the unit areas 411 is less than the anonymity threshold value 401, the de-

identification device may determine the de-identified areas 431 to 441 by grouping the arbitrary unit area 411 with a unit area adjacent to the arbitrary unit area 411.

When the number of pieces of location information included in the de-identified areas 431 to 441 is less than the anonymity threshold value 401, the de-identification
5 device may group the de-identified areas 431 to 441 with a unit area 411 adjacent to the de-identified areas 431 to 441.

That is, when the number of pieces of location information included in the grouped unit areas 411 is equal to or greater than the anonymity threshold value 401, the de-identification device may determine the grouped unit areas 411 as the de-identified
10 areas 431 to 441.

For example, the de-identification device may compare the number of pieces of location information included in each of the unit areas 411 to the anonymity threshold value 401 that is preset. For example, the de-identification device may first select an arbitrary unit area 411 and compare the number of pieces of location information included
15 in the arbitrary unit area 411 to the anonymity threshold value 401.

For example, the de-identification device may select the unit area 411 including the most pieces of location information among the unit areas 411 and compare the number of pieces of location information included in the selected unit area 411 to the anonymity threshold value 401.

20 When the number of pieces of location information included in the selected unit area 411 is greater than or equal to the anonymity threshold value 401, the de-identification device may determine the selected unit area 411 as the de-identified areas 431 to 441 without separate grouping.

For example, when the number of pieces of location information included in the
25 selected unit area 411 is less than the anonymity threshold value 401, the de-identification device may determine the unit areas 411 adjacent to the selected unit area 411, one unit area 411 among the de-identified areas 431 to 441, or the de-identified areas 431 to 441.

For example, when the number of pieces of location information included in the selected unit area 411 is less than the anonymity threshold value 401, the de-identification
30 device may determine the unit areas 411 adjacent to the selected unit area 411, the unit area 411 with the greatest number of pieces of location information among the de-identified areas 431 to 441, or the de-identified areas 431 to 441.

The de-identification device may generate new de-identified areas 431 to 441 by grouping the selected unit area 411 with the determined unit area 411 or the de-identification areas 431 to 441. The de-identification device may then combine the generated de-identified areas 431 to 441 with the adjacent unit areas 411 or the de-identified areas 431 to 441 until the number of pieces of location information included in the grouped unit areas 411 is greater than or equal to the anonymity threshold value 401.

For example, breadth-first search may be used in grouping the unit areas 411. For example, the de-identification device may, based on an arbitrary unit area 411, group the arbitrary unit area 411 with the adjacent unit areas 411 or the de-identified areas 431 to 441 until the number of pieces of location information is equal to or higher than the anonymity threshold value 401.

In determining the grouping of the unit area 411, a greedy algorithm may be used. That is, in a process of the grouping of the unit area 411, when the number of pieces of location information included in the grouped unit areas 411 is greater than or equal to the anonymity threshold value 401, the de-identification device may determine that the anonymity threshold value 401 is satisfied and determine the grouped unit areas 411 as the de-identified areas 431 to 441.

The de-identification device may then perform breadth-first search again on the arbitrary unit area 411 among the unit areas 411 that are not determined as the de-identified areas 431 to 441 to group the unit areas 411.

Accordingly, in the de-identified areas 431 to 441, the unit areas 411 may be grouped so that all of the unit areas 411 may belong to one of the de-identified areas 431 to 441. The number of pieces of location information included in each of the generated de-identified areas 431 to 441 may be greater than or equal to the anonymity threshold value 401. Referring to FIG. 4, an entire area 430 may be composed of the plurality of de-identified areas 431 to 441.

For example, the de-identification device may generate a graph that uses each unit area 411 as a vertex and an edge shared between the adjacent unit areas 411 as an edge. A value of each vertex may be determined according to the number of pieces of location information included in the unit area 411 corresponding to the vertex.

The de-identification device may, based on an arbitrary vertex in the generated graph, group vertices that satisfy the condition by performing breadth-first search until

the sum of the values of selected vertices is greater than the anonymity threshold value 401.

The de-identification device may generate the de-identified areas 431 to 441 so that the number of the de-identified areas 431 to 441 increases as the anonymity threshold value 401 decreases. That is, the number of the de-identified areas 431 to 441 that are
5 finally determined may increase as the anonymity threshold value 401 decreases.

The de-identification device may generate the de-identified areas 431 to 441 so that the number of the de-identified areas 431 to 441 decreases as the anonymity threshold value 401 increases. That is, the number of the de-identified areas 431 to 441 that are
10 finally determined may decrease as the anonymity threshold value 401 increases.

For example, in the grouping of the unit areas 411, different combinations of the de-identified areas 431 to 441 may be generated depending on a starting position of the unit area 411 that is initially selected, a direction of breadth-first search, and the like. That is, under the condition that the number of pieces of location information included in
15 all of the de-identified areas 431 to 441 is greater than or equal to the anonymity threshold value 401, various combinations of the de-identified areas 431 to 441 may be generated.

The de-identification device may generate a plurality of combinations of the de-identified areas 431 to 441 by grouping the unit areas 411 in a plurality of different ways and determine a combination of the de-identified areas 431 to 441 with the highest
20 usability evaluation result among the generated combinations of the de-identified areas 431 to 441 as a final de-identification result.

For example, a data usability index function may be used in a usability evaluation. For example, the de-identification device may determine a value of the data usability index function for each of the plurality of combinations of the de-identified areas 431 to
25 441 and determine a combination of the de-identified areas 431 to 441 with the greatest value of the data usability index function as a final de-identification result.

Alternatively, for example, the data usability index function may be used in a usability evaluation. For example, the de-identification device may determine a value of the data usability index function for each of the plurality of combinations of the de-identified areas 431 to
30 identified areas 431 to 441 and determine a combination of the de-identified areas 431 to 441 with the greatest argmax value of the value of the data usability index function as a final de-identification result.

Alternatively, the de-identification device may generate a plurality of combinations of the de-identified areas 431 to 441 by grouping the unit areas 411 in a plurality of different ways and determine a combination of the de-identified areas 431 to 441 with the number of the de-identified areas 431 to 441 less than or equal to a preset standard number among the generated combinations of the de-identified areas 431 to 441 as a final de-identification result.

The de-identification device may generate the de-identified areas 431 to 441 so that the number of the de-identified areas 431 to 441 is less than or equal to a preset standard number. That is, a resolution of de-identified data may be guaranteed due to the limitation of the number.

FIG. 5 is a diagram illustrating a result of de-identifying location information of users belonging to different categories, according to an embodiment.

For example, the users belonging to a category may refer to users belonging to a category in which some aspects of personal information are common, such as men in their 20s and women in their 30s. When location information of users for multiple categories is de-identified, de-identified areas 501 to 503 for each category may be determined. In addition, a de-identification device may generate a multi-dimensional de-identified area 504 by combining the de-identified areas 501 to 503 for each category.

According to an embodiment of the present disclosure, various types of personal information may be comprehensively used along with location information. A de-identification process for each category may be the same as the de-identification process described above. The multi-dimensional de-identified area 504 according to an embodiment of the present disclosure may be used for big data analytics for various services.

FIG. 6 is a flowchart illustrating a de-identification method according to an embodiment.

In operation 601, a de-identification device may identify an entire area including pieces of location information and an anonymity threshold value. In operation 602, the de-identification device may divide the entire area into a plurality of unit areas.

The de-identification device may divide the entire area into the unit areas without any gaps. However, the width or volume of all unit areas may not be the same. The de-identification device may divide the entire area into the unit areas so that at least one edge

or face is shared between adjacent unit areas. Depending on an embodiment, unit areas of various shapes may be used under the condition that at least one edge or face is shared between adjacent unit areas.

In operation 603, the de-identification device may determine the number of
5 pieces of location information included in each of the unit areas. In operation 604, the de-identification device may generate de-identification areas by grouping the unit areas based on the anonymity threshold value and the number of pieces of location information included in each of the unit areas.

When the number of pieces of location information included in an arbitrary unit
10 area among the unit areas is less than the anonymity threshold value, the de-identification device may determine the de-identified area by grouping the arbitrary unit area with a unit area adjacent to the arbitrary unit area.

When the number of pieces of location information included in the de-identified
15 area is less than the anonymity threshold value, the de-identification device may group the de-identified area with a unit area adjacent to the de-identified area.

That is, when the number of pieces of location information included in the grouped unit areas is equal to or higher than the anonymity threshold value, the de-identification device may determine the grouped unit areas as the de-identified area.

For example, the de-identification device may compare the number of pieces of
20 location information included in each of the unit areas to the anonymity threshold value that is preset. For example, the de-identification device may first select an arbitrary unit area and compare the number of pieces of location information included in the arbitrary unit area to the anonymity threshold value.

For example, the de-identification device may select a unit area including the
25 most pieces of location information among the unit areas and compare the number of pieces of location information included in the selected unit area to the anonymity threshold value. When the number of pieces of location information included in the selected unit area is greater than or equal to the anonymity threshold value, the de-identification device may determine the selected unit area as the de-identified area
30 without separate grouping.

The de-identification device may generate a plurality of combinations of de-identified areas by grouping the unit areas in a plurality of different ways and determine

a combination of de-identified areas with the highest usability evaluation result among the generated combinations of de-identified areas as a final de-identification result.

For example, a data usability index function may be used in a usability evaluation. For example, the de-identification device may determine a value of the data usability index function for each of the plurality of combinations of de-identified areas and
5 determine a combination of de-identified areas with the greatest value of the data usability index function as a final de-identification result.

The method according to embodiments may be written in a computer-executable program and may be implemented as various recording media such as magnetic storage
10 media, optical reading media, or digital storage media.

Various techniques described herein may be implemented in digital electronic circuitry, computer hardware, firmware, software, or combinations thereof. The implementations may be achieved as a computer program product, for example, a computer program tangibly embodied in a machine-readable storage device (a computer-
15 readable medium) to process the operations of a data processing device, for example, a programmable processor, a computer, or a plurality of computers or to control the operations. A computer program, such as the computer program(s) described above, may be written in any form of a programming language, including compiled or interpreted languages, and may be deployed in any form, including as a stand-alone program or as a
20 module, a component, a subroutine, or other units suitable for use in a computing environment. A computer program may be deployed to be processed on one computer or multiple computers at one site or distributed across multiple sites and interconnected by a communication network.

Processors suitable for processing of a computer program include, by way of
25 example, both general and special-purpose microprocessors, and any one or more processors of any type of digital computer. Generally, a processor will receive instructions and data from a read-only memory (ROM) or a random-access memory (RAM), or both. Elements of a computer may include at least one processor for executing instructions and one or more memory devices for storing instructions and data.
30 Generally, a computer also may include, or be operatively coupled to receive data from or transfer data to, or both, one or more mass storage devices for storing data, e.g., magnetic, magneto-optical disks, or optical disks. Examples of information carriers

suitable for embodying computer program instructions and data include semiconductor memory devices, for example, magnetic media such as hard disks, floppy disks, and magnetic tape, optical media such as compact disc ROM (CD-ROM) or a digital versatile disc (DVD), magneto-optical media such as floptical disks, ROM, RAM, flash memory, erasable programmable ROM (EPROM), or electrically erasable programmable ROM (EEPROM). The processor and the memory may be supplemented by, or incorporated in, special-purpose logic circuitry.

In addition, non-transitory computer-readable media may be any available media that may be accessed by a computer and may include both computer storage media and transmission media.

Although the present specification includes details of a plurality of specific embodiments, the details should not be construed as limiting any invention or a scope that can be claimed, but rather should be construed as being descriptions of features that may be peculiar to specific embodiments of specific inventions. Specific features described in the present specification in the context of individual embodiments may be combined and implemented in a single embodiment. On the contrary, various features described in the context of a single embodiment may be implemented in a plurality of embodiments individually or in any appropriate sub-combination. Moreover, although features may be described above as acting in specific combinations and even initially claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claimed combination may be changed to a sub-combination or a modification of a sub-combination.

Likewise, although operations are depicted in a predetermined order in the drawings, it should not be construed that the operations need to be performed sequentially or in the predetermined order, which is illustrated to obtain a desirable result, or that all of the shown operations need to be performed. In specific cases, multi-tasking and parallel processing may be advantageous. In addition, it should not be construed that the separation of various device components of the aforementioned embodiments is required in all types of embodiments, and it should be understood that the described program components and devices are generally integrated as a single software product or packaged into a multiple-software product.

The embodiments disclosed in the present specification and the drawings are

intended merely to present specific examples in order to aid in understanding of the present disclosure, but are not intended to limit the scope of the present disclosure. It will be apparent to one of ordinary skill in the art that, in addition to the disclosed embodiments, various other examples modified based on the technical spirit of the present disclosure may be implemented.

5

CLAIMS

1. A method of de-identifying location information, the method comprising:
identifying an anonymity threshold value and an entire area comprising pieces of
location information of users;
5 dividing the entire area into a plurality of unit areas;
determining a number of pieces of location information included in each of the
unit areas; and
generating de-identified areas by grouping the unit areas based on the anonymity
threshold value and the number of pieces of location information included in each of the
10 unit areas.
2. The method of claim 1, wherein a number of the generated de-identified
areas increases as the anonymity threshold value decreases, and decreases as the
anonymity threshold value increases.
15
3. The method of claim 1, wherein the generating of the de-identified areas
comprises:
determining, when a number of pieces of location information included in an
arbitrary unit area among the unit areas is less than the anonymity threshold value, the
20 de-identified areas by grouping the arbitrary unit area with a unit area adjacent to the
arbitrary unit area.
4. The method of claim 1, wherein the generating of the de-identified areas
comprises:
25 grouping, when a number of pieces of location information included in a de-
identified area is less than the anonymity threshold value, the de-identified area with a
unit area adjacent to the de-identified area.
5. The method of claim 1, wherein a number of pieces of location
30 information included in each of the generated de-identified areas is greater than or equal
to the anonymity threshold value.

6. The method of claim 5, wherein the generating of the de-identified areas further comprises:

generating a plurality of combinations of de-identified areas by grouping the unit areas in a plurality of different ways; and

5 determining a combination of de-identified areas with a highest usability evaluation result among the generated combinations of de-identified areas as a final de-identification result.

7. The method of claim 5, wherein the generating of the de-identified areas
10 comprises:

generating a plurality of combinations of de-identified areas by grouping the unit areas in a plurality of different ways; and

determining a combination of de-identified areas with a number of the de-identified areas less than or equal to a preset standard number among the generated
15 combinations of de-identified areas as a final de-identification result.

8. A method of de-identifying location information, the method comprising:
identifying a geographic area comprising pieces of location information of users
belonging to an arbitrary category;

20 dividing the geographic area into a plurality of unit areas for each dimension;

comparing a number of pieces of location information included in each of the unit areas to an anonymity threshold value that is preset;

generating de-identified areas by grouping the unit areas based on the comparison result; and

25 obtaining a multi-dimensional de-identified area by combining i) a de-identified area with ii) a de-identified area generated from a geographic area comprising pieces of location information of users belonging to a category other than the arbitrary category.

9. The method of claim 8, wherein a number of the generated de-identified
30 areas increases as the anonymity threshold value decreases, and decreases as the anonymity threshold value increases.

10. The method of claim 8, wherein the generating of the de-identified areas comprises:

determining, when a number of pieces of location information included in an arbitrary unit area among the unit areas is less than the anonymity threshold value, the
5 de-identified areas by grouping the arbitrary unit area with a unit area adjacent to the arbitrary unit area.

11. The method of claim 8, wherein the generating of the de-identified areas comprises:

10 grouping, when a number of pieces of location information included in a de-identified area is less than the anonymity threshold value, the de-identified area with a unit area adjacent to the de-identified area.

12. The method of claim 8, wherein a number of pieces of location
15 information included in each of the generated de-identified areas is greater than or equal to the anonymity threshold value.

13. A device for de-identifying location information, the device comprising:
a processor,

20 wherein the processor is configured to:

identify an anonymity threshold value and an entire area comprising
pieces of location information of users;

divide the entire area into a plurality of unit areas;

25 determine a number of pieces of location information included in each
of the unit areas; and

generate de-identified areas by grouping the unit areas based on the
anonymity threshold value and the number of pieces of location information
included in each of the unit areas.

30 14. The device of claim 13, wherein a number of the generated de-identified
areas increases as the anonymity threshold value decreases, and decreases as the
anonymity threshold value increases.

15. The device of claim 13, wherein the processor is configured to:
determine, when a number of pieces of location information included in an
arbitrary unit area among the unit areas is less than the anonymity threshold value, the
5 de-identified areas by grouping the arbitrary unit area with a unit area adjacent to the
arbitrary unit area.

16. The device of claim 13, wherein the processor is configured to:
group, when a number of pieces of location information included in a de-
10 identified area is less than the anonymity threshold value, the de-identified area with a
unit area adjacent to the de-identified area.

17. The device of claim 13, wherein a number of pieces of location
information included in each of the generated de-identified areas is greater than or equal
15 to the anonymity threshold value.

18. The device of claim 17, wherein the processor is configured to:
generate a plurality of combinations of de-identified areas by grouping the unit
areas in a plurality of different ways; and
20 determine a combination of de-identified areas with a highest usability evaluation
result among the generated combinations of de-identified areas as a final de-identification
result.

19. The device of claim 17, wherein the processor is configured to:
25 generate a plurality of combinations of de-identified areas by grouping the unit
areas in a plurality of different ways; and
determine a combination of de-identified areas with a number of the de-identified
areas less than or equal to a preset standard number among the generated combinations
of de-identified areas as a final de-identification result.

30

20. A device for de-identifying location information, the device comprising:
a processor,

wherein the processor is configured to:

identify a geographic area comprising pieces of location information of users belonging to an arbitrary category;

5 divide the geographic area into a plurality of unit areas for each dimension;

compare a number of pieces of location information included in each of the unit areas to an anonymity threshold value that is preset;

generate de-identified areas by grouping the unit areas based on the comparison result; and

10 obtain a multi-dimensional de-identified area by combining i) a de-identified area with ii) a de-identified area generated from a geographic area comprising pieces of location information of users belonging to a category other than the arbitrary category.