

NATURAL-LANGUAGE DATA QUERY INTENTION DETERMINING METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

5 The present application claims priority to the Chinese Patent Application No. CN 201811021831.1, filed on August 31, 2018 and entitled "NATURAL-LANGUAGE DATA QUERY INTENTION DETERMINING METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM", which is incorporated herein by reference in its entity.

10 **TECHNICAL FIELD**

 The present application relates to the field of data query technologies, and in particular, to a natural-language data query intention determining method and apparatus, a computer device, and a storage medium.

15 **BACKGROUND**

 At present, when querying data, a user usually needs to select options on a query interface based on the user's intention. Subsequently, a server forms query content of the user from the query items selected by the user, and obtains a query result based on the query content.

 However, because a search is performed through query item selection, the user can only select
20 provided selection items. When there are few selection items, the user's selection range is limited. When there are excessive selection items, the user's selection operation is relatively complicated. Natural language-based queries become important in the future, to facilitate searches and queries. Identifying a natural language-based data query intention is the basis for natural language-based queries.

25 Therefore, a natural-language data query intention determining method and apparatus, a computer device, and a storage medium need to be provided, to accurately capture a user's data

query intention from a natural language and clearly extract a question range and an analysis dimension.

SUMMARY

5 An objective of the present application is to provide a natural-language data query intention determining method and apparatus, a computer device, and a storage medium, to resolve the above-mentioned problems in the prior art.

To achieve the foregoing objective, the present application provides a natural-language data query intention determining method.

10 The method includes: presetting a plurality of keyword filters representing a query range, where each of the keyword filters corresponds to a range word set, and the range word set includes a plurality of range words; obtaining a natural language-based data query request to be analyzed; performing word segmentation on the data query request to obtain a first word set; obtaining a word in the first word set that is consistent with the range word through filtering by using each of the
15 keyword filters; obtaining a query range of the data query request based on all words obtained through filtering; filtering out the word in the first word set that is consistent with the range word to obtain a second word set; performing semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result; determining a data presentation dimension and a data presentation method corresponding to the data query request
20 based on the semantic analysis result; and outputting the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

To achieve the foregoing objective, the present application provides a natural-language data query intention determining apparatus.

25 The apparatus includes: a plurality of keyword filters, where the keyword filters are configured to represent a query range, each of the keyword filters corresponds to a range word set, and the range word set includes a plurality of range words; an obtaining module, configured to obtain a natural language-based data query request to be analyzed; a word segmentation module,

configured to perform word segmentation on the data query request to obtain a first word set; an invoking module, configured to invoke each of the keyword filters to obtain a word in the first word set that is consistent with the range word through filtering; a first determining module, configured to obtain a query range of the data query request based on all words obtained through
 5 filtering; a filtering-out module, configured to filter out the word in the first word set that is consistent with the range word to obtain a second word set; an annotation module, configured to perform semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result; a second determining module, configured to determine a data presentation dimension and a data presentation method corresponding to the data query request
 10 based on the semantic analysis result; and an output module, configured to output the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

For the foregoing objective, the present application further provides a computer device including a memory, a processor, and a computer program that is stored in the memory and can
 15 operate on the processor, where when executing the program, the processor implements the following steps of a natural-language data query intention determining method:

presetting a plurality of keyword filters representing a query range, where each of the keyword filters corresponds to a range word set, and the range word set includes a plurality of range words; obtaining a natural language-based data query request to be analyzed; performing word
 20 segmentation on the data query request to obtain a first word set; obtaining a word in the first word set that is consistent with the range word through filtering by using each of the keyword filters; obtaining a query range of the data query request based on all words obtained through filtering; filtering out the word in the first word set that is consistent with the range word to obtain a second word set; performing semantic annotation on a word in the second word set based on a semantic
 25 knowledge base to generate a semantic analysis result; determining a data presentation dimension and a data presentation method corresponding to the data query request based on the semantic analysis result; and outputting the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

For the foregoing objective, the present application further provides a computer readable
 30 storage medium on which a computer program is stored, where when the program is executed by

a processor, the following steps of a natural-language data query intention determining method are implemented:

presetting a plurality of keyword filters representing a query range, where each of the keyword filters corresponds to a range word set, and the range word set includes a plurality of range words;

5 obtaining a natural language-based data query request to be analyzed; performing word segmentation on the data query request to obtain a first word set; obtaining a word in the first word set that is consistent with the range word through filtering by using each of the keyword filters; obtaining a query range of the data query request based on all words obtained through filtering; filtering out the word in the first word set that is consistent with the range word to obtain a second

10 word set; performing semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result; determining a data presentation dimension and a data presentation method corresponding to the data query request based on the semantic analysis result; and outputting the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

15 According to the natural-language data query intention determining method and apparatus, the computer device, and the storage medium provided in the present application, the keyword filters are preset, and range words in a range word set corresponding to a keyword filter can be obtained through filtering by using the keyword filter. Therefore, a range word combination may be set as required to filter predetermined range words. After obtaining the natural language-based

20 data query request, word segmentation is first performed on the data query request to obtain a word set. Then, the word set is filtered by using the preset keyword filters to obtain a word in the word set that is consistent with the range word, to form a query range based on the words obtained through filtering. After the query range is obtained, the words obtained through filtering in the word set are filtered out, semantic annotation is performed on a remaining word, and the data presentation

25 dimension and the data presentation method are determined based on a semantic annotation result. Finally, the query range, the data presentation dimension, and the data presentation method are used as the standardized data query intention corresponding to the data query request, so that a data query can be performed by using unified query logic based on the standardized query intention and query efficiency can be improved.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 is a flowchart of a natural-language data query intention determining method according to Embodiment 1 of the present application;

FIG. 2 is a block diagram of a natural-language data query intention determining apparatus
5 according to Embodiment 2 of the present application; and

FIG. 3 is a diagram of a hardware structure of a computer device according to Embodiment 3 of the present application.

DETAILED DESCRIPTION

10 To make the objectives, technical solutions, and advantages of the present application clearer and more comprehensible, the following further describes the present application in detail with reference to the accompanying drawings and embodiments. It should be understood that the specific embodiments described herein are merely intended to explain the present application but not limit the present application. Based on the embodiments of the present application, all other
15 embodiments obtained by a person of ordinary skill in the art without creative efforts shall fall within the protection scope of the present application.

Embodiment 1

Embodiment 1 provides a natural-language data query intention determining method, so that
20 a user can query content based on natural language expressions to improve query efficiency and a server can directly generate a query result based on the user's natural language. The method converts a natural language-based data query request into a standardized data query intention. The data query intention includes a query range, a data presentation dimension, and a data presentation method. Specifically, FIG. 1 is a flowchart of the natural-language data query intention determining
25 method provided by Embodiment 1 of the present application. As shown in FIG. 1, the natural-language data query intention determining method includes the following steps S101 to S109.

Step S101: Preset a plurality of keyword filters representing a query range.

Each of the keyword filters corresponds to a range word set, and the range word set includes a plurality of range words. For example, a range word set corresponding to one keyword filter is set to include Shanghai, Tianjin, Beijing, Nanchang, Shenyang, and the like. For another example, a range word set corresponding to another keyword filter includes old-age, infant, teenager, middle-age, and the like. A range word set corresponding to a keyword filter can be specifically set based on data content.

Step S102: Obtain a natural language-based data query request to be analyzed.

A user query request receiving interface is provided to receive the natural language-based data query request that is entered by a user. Specifically, a data query request may be received through text input or voice input. The voice input may be identified as text at a back end. Regardless of an input manner, the natural language-based data query request is directly embodied through a natural language. For example, the natural language-based data query request is "what is the overall overdue status of customers in the Shanghai district?" For another example, the natural language-based data query request may be "what were the total sales of the first sales group in March?"

Step S103: Perform word segmentation on the data query request to obtain a first word set.

For example, word segmentation is performed on a data query request "how are the education backgrounds of male users in the Shanghai district distributed?" to obtain the first word set: "how", "are", "the", "education", "backgrounds", "of", "male", "users", "in", "the", "Shanghai", "district", and "distributed". Preferably, useless words can be filtered out from the first word set, such as, "are", "of", and the like.

Step S104: Obtain a word in the first word set that is consistent with the range word through filtering by using each of the keyword filters.

In the present step, words in the first word set are filtered by using each preset keyword filter to obtain the word in the first word set that is consistent with the range word.

Step S105: Obtain a query range of the data query request based on all words obtained through filtering.

In step S104, one word consistent with the range word may be obtained by using one filter, or a plurality of words consistent with the range words may be obtained by using a plurality of filters.

When a plurality of words consistent with the range words are obtained by using a plurality of filters, all the words consistent with the range words are combined to obtain the query range of the data query request. For example, when the first word set corresponding to "how are the education backgrounds of male users in the Shanghai district distributed?" is filtered by using a keyword filter, obtained words consistent with the range words include "Shanghai" and "male". Therefore, a query range of the data query request is "Shanghai male".

Step S106: Filter out the word in the first word set that is consistent with the range word to obtain a second word set.

When the words consistent with the range words and obtained in step S104 include "Shanghai" and "male", a second word set obtained after these two words are filtered out include "district", "users", "education", "backgrounds", "how", and "distributed".

Step S107: Perform semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result.

For example, the semantic analysis result includes a plurality of word-semantics pairs, and the word-semantics pair includes one word in the second word set and semantics of the word.

Semantics of a word includes a part of speech and a meaning of the word. When semantic annotation is performed, parameters corresponding to each word are determined based on the semantic knowledge base. For example, the words in the second word set include "district", "users", "education", "backgrounds", "how", and "distributed". Based on the semantic knowledge base, it can be determined that "district" is a noun and an administrative division unit, "users" is a noun and a human being category, "education" and "backgrounds" are nouns and denote a way to describe a degree of education, "how" is a question word, and "distributed" is a verb. In this case, a semantic analysis result is presented as follows:

<district, noun, an administrative division unit>;

<users, noun, a human being category>;

<education, backgrounds, nouns, a way to describe a degree of education>;

<how, question word, denoting a question>; and

<distributed, verb, denoting that a certain object is spread in a specific range>.

Step S108: Determine a data presentation dimension and a data presentation method corresponding to the data query request based on the semantic analysis result.

For determining of a data presentation dimension, semantics corresponding to each data presentation dimension in the data content are preset. For example, the data content is census data for nationwide young people from 25 to 30 years old. Data presentation dimensions include "education background", "annual income", "annual consumption", "fertility", and the like. Semantics corresponding to "education background", "annual income", "annual consumption", and "fertility" are preset.

In the present step, the determining a data presentation dimension corresponding to the data query request based on the semantic analysis result specifically includes:

matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation dimension in the data content; and

when the semantics in the word-semantics pair are the same as semantics corresponding to a first data presentation dimension in the data content, using the first data presentation dimension as the data presentation dimension corresponding to the data query request.

The data presentation method refers to a way of data expression. For example, in terms of age data for students of a school, data presentation methods include presenting data by averages of all ages, by proportions of all age brackets, by distribution of all ages, and the like.

For determining of a data presentation method, semantics corresponding to each data presentation method in the data content are preset. For example, the data content is census data for nationwide young people from 25 to 30 years old. Data presentation methods include "distribution", "average", and the like. Semantics corresponding to "distribution" and "average" are preset. In the present step, the determining a data presentation method corresponding to the data query request based on the semantic analysis result specifically includes:

matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation method in the data content; and

when the semantics in the word-semantics pair are the same as semantics corresponding to a

first data presentation method in the data content, using the first data presentation method as the data presentation method corresponding to the data query request.

Step S109: Output the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

- 5 Each output data query intention includes a query range, a data presentation dimension, and a data presentation method and is a standardized and unified data query intention.

According to the natural-language data query intention determining method provided in this embodiment, the keyword filters are preset, and range words in a range word set corresponding to a keyword filter can be obtained through filtering by using the keyword filter. Therefore, a range
 10 word combination may be set as required to filter predetermined range words. After obtaining the natural language-based data query request, word segmentation is first performed on the data query request to obtain a word set. Then, the word set is filtered by using the preset keyword filters to obtain a word in the word set that is consistent with the range word, to form a query range based on the words obtained through filtering. After the query range is obtained, the words obtained
 15 through filtering in the word set are filtered out, semantic annotation is performed on a remaining word, and the data presentation dimension and the data presentation method are determined based on a semantic annotation result. Finally, the query range, the data presentation dimension, and the data presentation method are used as the standardized data query intention corresponding to the data query request, so that a data query can be performed by using unified query logic based on the
 20 standardized query intention and query efficiency can be improved.

In general, one word consistent with the range word can be obtained at most for one filter. In a specific embodiment, if two or more words consistent with the range words are obtained, a corresponding number of data query intentions can be obtained by using the data query request. Each data query intention has a different query range but the same data presentation dimension and
 25 data presentation method. For example, the data query request is "how are the education backgrounds of male users in the Shanghai and Beijing districts distributed?", and obtained query ranges include "male in the Shanghai district" and "male in the Beijing district".

Optionally, a data query intention of an incomplete data query request may be determined, that is, a data query intention may be determined for a data query request that includes partial

content of a query range, a data presentation dimension, and a data presentation method. In this embodiment provided in the present application, the incomplete data query request is first supplemented, and then the data query intention is determined, so that a standardized and unified data query intention can be output for the incomplete data query request.

5 Specifically, in an embodiment, historical data query intentions are obtained when each of the keyword filters fails to obtain the word in the first word set that is consistent with the range word through filtering. A historical data query intention that has a highest degree of matching with the data presentation dimension and the data presentation method is determined. Specifically, a plurality of entries of historical data query intentions are searched for a historical data query
10 intention that includes the data presentation dimension and the data presentation method. If one entry of historical data query intention is found, this entry of historical data query intention is the historical data query intention that has the highest degree of matching with the data presentation dimension and the data presentation method. If a plurality of entries of historical data query intentions are found and the found entries of historical data query intentions include historical data
15 query intentions with the same content, a historical data query intention with the same content that has the most entries is used as the historical data query intention with the highest degree of matching with the data presentation dimension and the data presentation method. A historical query range in the historical data query intention with the highest degree of matching is used as the query range of the data query request. For example, if a data query request to be analyzed is "average
20 annual income", a data presentation dimension is "annual income", and a data presentation method is "average". In this case, a plurality of entries of historical data query intentions are searched for a historical data query intention including the most entries of "annual income" and "average", and this found historical data query intention is used as the historical data query intention with the highest degree of matching. A historical query range in the historical data query intention with the
25 highest degree of matching is "Beijing young people", and "Beijing young people" is used as a query range corresponding to the data query request "average annual income". A text font color representing the query range is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request, that is, a text font color of "Beijing young people" is set to gray when a
30 standardized data query intention is output for the data query request "average annual income", to specially remind the user, so that the user can determine the query range.

In another embodiment, historical data query intentions are obtained when the data presentation dimension corresponding to the data query request cannot be determined based on the semantic analysis result. A historical data query intention that has a highest degree of matching with the query range and the data presentation method is determined. Specifically, a plurality of entries of historical data query intentions are searched for a historical data query intention that includes the query range and the data presentation method. If one entry of historical data query intention is found, this entry of historical data query intention is the historical data query intention that has the highest degree of matching with the query range and the data presentation method. If a plurality of entries of historical data query intentions are found and the found entries of historical data query intentions include historical data query intentions with the same content, a historical data query intention with the same content that has the most entries is used as the historical data query intention with the highest degree of matching with the query range and the data presentation method. A historical data presentation dimension in the historical data query intention with the highest degree of matching is used as the data presentation dimension of the data query request.

For example, if a data query request to be analyzed is "Shanghai male average status", a query range is "Shanghai male", and a data presentation method is "average". In this case, a plurality of entries of historical data query intentions are searched for a historical data query intention including the most entries of "Shanghai male" and "average", and this found historical data query intention is used as the historical data query intention with the highest degree of matching. A data presentation dimension in the historical data query intention with the highest degree of matching is "annual income", and "annual income" is used as a data presentation dimension corresponding to the data query request "Shanghai male average status". A text font color representing the data presentation dimension is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request, that is, a text font color of "annual income" is set to gray when a standardized data query intention is output for the data query request "Shanghai male average status", to specially remind the user, so that the user can determine the data presentation dimension.

In still another embodiment, historical data query intentions are obtained when the data presentation method corresponding to the data query request cannot be determined based on the semantic analysis result. A historical data query intention that has a highest degree of matching with the query range and the data presentation dimension is determined. Specifically, a plurality of

entries of historical data query intentions are searched for a historical data query intention that includes the query range and the data presentation dimension. If one entry of historical data query intention is found, this entry of historical data query intention is the historical data query intention that has the highest degree of matching with the query range and the data presentation dimension.

5 If a plurality of entries of historical data query intentions are found and the found entries of historical data query intentions include historical data query intentions with the same content, a historical data query intention with the same content that has the most entries is used as the historical data query intention with the highest degree of matching with the query range and the data presentation dimension. A historical data presentation method in the historical data query
10 intention with the highest degree of matching is used as the data presentation method of the data query request. For example, if a data query request to be analyzed is "Shanghai male annual income", a query range is "Shanghai male", and a data presentation dimension is "annual income". In this case, a plurality of entries of historical data query intentions are searched for a historical data query intention including the most entries of "Shanghai male" and "annual income", and this
15 found historical data query intention is used as the historical data query intention with the highest degree of matching. A data presentation method in the historical data query intention with the highest degree of matching is "average", and "average" is used as a data presentation method corresponding to the data query request "Shanghai male annual income". A text font color representing the data presentation method is set to gray when the query range, the data presentation
20 dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request, that is, a text font color of "average" is set to gray when a standardized data query intention is output for the data query request "Shanghai male annual income", to specially remind the user, so that the user can determine the data presentation method.

25 **Embodiment 2**

Corresponding to Embodiment 1, Embodiment 2 of the present application provides a natural-language data query intention determining apparatus. For related parts, mutual reference may be made to those in Embodiment 1. FIG. 2 is a block diagram of a natural-language data query intention determining apparatus according to Embodiment 2 of the present application. As shown
30 in FIG. 2, the apparatus includes a plurality of keyword filters 201, an obtaining module 202, a

word segmentation module 203, an invoking module 204, a first determining module 205, a filtering-out module 206, an annotation module 207, a second determining module 208, and an output module 209.

The keyword filters 201 are configured to represent a query range, where each of the keyword filters corresponds to a range word set, and the range word set includes a plurality of range words; the obtaining module 202 is configured to obtain a natural language-based data query request to be analyzed; the word segmentation module 203 is configured to perform word segmentation on the data query request to obtain a first word set; the invoking module 204 is configured to invoke each of the keyword filters to obtain a word in the first word set that is consistent with the range word through filtering; the first determining module 205 is configured to obtain a query range of the data query request based on all words obtained through filtering; the filtering-out module 206 is configured to filter out the word in the first word set that is consistent with the range word to obtain a second word set; the annotation module 207 is configured to perform semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result; the second determining module 208 is configured to determine a data presentation dimension and a data presentation method corresponding to the data query request based on the semantic analysis result; and; the output module 209 is configured to output the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

According to the natural-language data query intention determining apparatus provided in this embodiment, the keyword filters are preset, and range words in a range word set corresponding to a keyword filter can be obtained through filtering by using the keyword filter. Therefore, a range word combination may be set as required to filter predetermined range words. After the obtaining module obtains the natural language-based data query request, the word segmentation module first performs word segmentation on the data query request to obtain a word set. Then, the invoking module invokes the preset keyword filters to filter the word set, to obtain a word in the word set that is consistent with the range word. The first determining module can form a query range based on the words obtained through filtering. After the query range is obtained, the filtering-out module filters out the words obtained through filtering in the word set, the annotation module performs semantic annotation on a remaining word, and the second determining module determines the data

presentation dimension and the data presentation method based on a semantic annotation result. Finally, the output module uses the query range, the data presentation dimension, and the data presentation method as the standardized data query intention corresponding to the data query request, so that a data query can be performed by using unified query logic based on the standardized query intention and query efficiency can be improved.

Preferably, the apparatus further includes a first supplement module. When each of the keyword filters fails to obtain the word in the first word set that is consistent with the range word through filtering, the first supplement module is configured to perform the following steps: obtaining historical data query intentions; determining a historical data query intention that has a highest degree of matching with the data presentation dimension and the data presentation method; and obtaining a historical query range in the historical data query intention with the highest degree of matching as the query range of the data query request. A text font color representing the query range is set to gray when the output module outputs the query range, the data presentation dimension, and the data presentation method as the standardized data query intention corresponding to the data query request.

Preferably, the apparatus further includes a second supplement module. When the second determining module cannot determine the data presentation dimension corresponding to the data query request based on the semantic analysis result, the second supplement module is configured to perform the following steps: obtaining historical data query intentions; determining a historical data query intention that has a highest degree of matching with the query range and the data presentation method; and obtaining a historical data presentation dimension in the historical data query intention with the highest degree of matching as the data presentation dimension of the data query request. A text font color representing the data presentation dimension is set to gray when the output module outputs the query range, the data presentation dimension, and the data presentation method as the standardized data query intention corresponding to the data query request.

Preferably, the apparatus further includes a third supplement module. When the second determining module cannot determine the data presentation method corresponding to the data query request based on the semantic analysis result, the third supplemental module is configured to perform the following steps: obtaining historical data query intentions; determining a historical

data query intention that has a highest degree of matching with the query range and the data presentation dimension; and obtaining a historical data presentation method in the historical data query intention with the highest degree of matching as the data presentation method of the data query request; where a text font color representing the data presentation method is set to gray when the output module outputs the query range, the data presentation dimension, and the data presentation method as the standardized data query intention corresponding to the data query request.

Preferably, the semantic analysis result generated by the annotation module includes a plurality of word-semantics pairs, and the word-semantics pair includes one word in the second word set and semantics of the word.

The apparatus further includes a first presetting module, configured to preset semantics corresponding to each data presentation dimension in data content. When determining the data presentation dimension corresponding to the data query request based on the semantic analysis result, the second determining module specifically performs the following steps: matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation dimension in the data content; and when the semantics in the word-semantics pair are the same as semantics corresponding to a first data presentation dimension in the data content, using the first data presentation dimension as the data presentation dimension corresponding to the data query request.

The apparatus further includes a second presetting module, configured to preset semantics corresponding to each data presentation method in data content. When determining the data presentation method corresponding to the data query request based on the semantic analysis result, the second determining module specifically performs the following steps: matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation method in the data content; and when the semantics in the word-semantics pair are the same as semantics corresponding to a first data presentation method in the data content, using the first data presentation method as the data presentation method corresponding to the data query request.

Embodiment 3

This embodiment further provides a computer device, for example, a smartphone, a tablet computer, a notebook computer, a desktop computer, a rack server, a blade server, a tower server, or a cabinet server (including an independent server or a server cluster composed of a plurality of servers) that can execute a program. As shown in FIG. 3, a computer device 20 in this embodiment includes at least but is not limited to a memory 21 and a processor 22 that can communicate with and connect to each other through a system bus. It should be noted that FIG. 3 shows only the computer device 20 with the components 21 and 22, but it should be understood that not all shown components need to be implemented and more or fewer components can be implemented instead.

In this embodiment, the memory 21 (namely, a readable storage medium) includes a flash memory, a hard disk, a multimedia card, a card type memory (such as an SD or DX memory), a random access memory (RAM), a static random access memory (SRAM), a read-only memory (ROM), an electrically erasable programmable read-only memory (EEPROM), a programmable read-only memory (PROM), a magnetic memory, a magnetic disk, an optical disc, and the like. In some embodiments, the memory 21 may be an internal storage unit of the computer device 20, such as a hard disk or a memory of the computer device 20. In other embodiments, the memory 21 may alternatively be an external storage device of the computer device 20, such as a plug-in hard disk equipped on the computer device 20, a smart media card (Smart Media Card, SMC), a secure digital (Secure Digital, SD) card, a flash card (Flash Card), or the like. Certainly, the memory 21 may alternatively include both an internal storage unit and an external storage device of the computer device 20. In this embodiment, the memory 21 is generally configured to store an operating system and various application software installed in the computer device 20, such as program code of the natural-language data query intention determining apparatus according to Embodiment 2. In addition, the memory 21 may further be configured to temporarily store various types of data that has been or is to be output.

In some embodiments, the processor 22 may be a central processing unit (Central Processing Unit, CPU), a controller, a microcontroller, a microprocessor, or another data processing chip. The processor 22 is typically configured to control overall operations of the computer device 20. In this embodiment, the processor 22 is configured to run the program code stored in the memory 21 or process data, such as the natural-language data query intention determining apparatus.

Embodiment 4

This embodiment further provides a computer readable storage medium, such as a flash memory, a hard disk, a multimedia card, a card type memory (such as an SD or DX memory), a random access memory (RAM), a static random access memory (SRAM), a read-only memory (ROM), an electrically erasable programmable read-only memory (EEPROM), a programmable
5 read-only memory (PROM), a magnetic memory, a magnetic disk, an optical disc, a server, an App store, or the like. The computer readable storage medium stores a computer program. When the program is executed by a processor, a corresponding function is implemented. The computer readable storage medium according to this embodiment is used in the natural-language data query intention determining apparatus. When the computer readable storage medium is executed by the
10 processor, the natural-language data query intention determining method according to Embodiment 1 is implemented.

Sequence numbers of the embodiments of the present application are merely for a purpose of description, and do not represent advantages or disadvantages of the embodiments.

From the descriptions of the implementations, a person skilled in the art can clearly understand
15 that the method in the embodiments may be implemented by software and a necessary universal hardware platform, and certainly may alternatively be implemented by hardware. However, in many cases, the implementation performed by software and a necessary universal hardware platform is better.

The foregoing descriptions are only preferred embodiments of the present application, and are
20 not intended to limit the scope of the present application. Any equivalent structure or equivalent process transformation that is made by using the content of the specification and accompanying drawings of the present application, or that is directly or indirectly applied in other related technical fields, is included within the protection scope of the present application.

CLAIMS

1. What is claimed is:

A natural-language data query intention determining method, comprising:

presetting a plurality of keyword filters representing a query range, wherein each of the
5 keyword filters corresponds to a range word set, and the range word set comprises a plurality of
range words;

obtaining a natural language-based data query request to be analyzed;

performing word segmentation on the data query request to obtain a first word set;

obtaining a word in the first word set that is consistent with the range word through filtering
10 by using each of the keyword filters;

obtaining a query range of the data query request based on all words obtained through filtering;

filtering out the word in the first word set that is consistent with the range word to obtain a
second word set;

performing semantic annotation on a word in the second word set based on a semantic
15 knowledge base to generate a semantic analysis result;

determining a data presentation dimension and a data presentation method corresponding to
the data query request based on the semantic analysis result; and

outputting the query range, the data presentation dimension, and the data presentation method
as a standardized data query intention corresponding to the data query request.

20 2. The natural-language data query intention determining method according to claim 1,
wherein the method further comprises:

obtaining historical data query intentions when each of the keyword filters fails to obtain the
word in the first word set that is consistent with the range word through filtering;

determining a historical data query intention that has a highest degree of matching with the
25 data presentation dimension and the data presentation method; and

obtaining a historical query range in the historical data query intention with the highest degree of matching as the query range of the data query request; wherein

a text font color representing the query range is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

3. The natural-language data query intention determining method according to claim 1, wherein the method further comprises:

obtaining historical data query intentions when the data presentation dimension corresponding to the data query request cannot be determined based on the semantic analysis result;

10 determining a historical data query intention that has a highest degree of matching with the query range and the data presentation method; and

obtaining a historical data presentation dimension in the historical data query intention with the highest degree of matching as the data presentation dimension of the data query request; wherein

15 a text font color representing the data presentation dimension is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

4. The natural-language data query intention determining method according to claim 1, wherein the method further comprises:

20 obtaining historical data query intentions when the data presentation method corresponding to the data query request cannot be determined based on the semantic analysis result;

determining a historical data query intention that has a highest degree of matching with the query range and the data presentation dimension; and

25 obtaining a historical data presentation method in the historical data query intention with the highest degree of matching as the data presentation method of the data query request; wherein

a text font color representing the data presentation method is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized

data query intention corresponding to the data query request.

5. The natural-language data query intention determining method according to claim 1, wherein the semantic analysis result comprises a plurality of word-semantics pairs, and the word-semantics pair comprises one word in the second word set and semantics of the word.

5 6. The natural-language data query intention determining method according to claim 5, wherein the method further comprises:

presetting semantics corresponding to each data presentation dimension in data content; and

the determining a data presentation dimension corresponding to the data query request based on the semantic analysis result comprises:

10 matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation dimension in the data content; and

when the semantics in the word-semantics pair are the same as semantics corresponding to a first data presentation dimension in the data content, using the first data presentation dimension as the data presentation dimension corresponding to the data query request.

15 7. The natural-language data query intention determining method according to claim 5, wherein the method further comprises:

presetting semantics corresponding to each data presentation method in data content;

the determining a data presentation method corresponding to the data query request based on the semantic analysis result comprises:

20 matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation method in the data content; and

when the semantics in the word-semantics pair are the same as semantics corresponding to a first data presentation method in the data content, using the first data presentation method as the data presentation method corresponding to the data query request.

25 8. A natural-language data query intention determining apparatus, comprising:

a plurality of keyword filters, wherein the keyword filters are configured to represent a query

range, each of the keyword filters corresponds to a range word set, and the range word set comprises a plurality of range words;

an obtaining module, configured to obtain a natural language-based data query request to be analyzed;

5 a word segmentation module, configured to perform word segmentation on the data query request to obtain a first word set;

an invoking module, configured to invoke each of the keyword filters to obtain a word in the first word set that is consistent with the range word through filtering;

10 a first determining module, configured to obtain a query range of the data query request based on all words obtained through filtering;

a filtering-out module, configured to filter out the word in the first word set that is consistent with the range word to obtain a second word set;

an annotation module, configured to perform semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result;

15 a second determining module, configured to determine a data presentation dimension and a data presentation method corresponding to the data query request based on the semantic analysis result; and

20 an output module, configured to output the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

9. A computer device, comprising a memory, a processor, and a computer program that is stored in the memory and can operate on the processor, wherein when executing the program, the processor implements the following steps of a natural-language data query intention determining method:

25 presetting a plurality of keyword filters representing a query range, wherein each of the keyword filters corresponds to a range word set, and the range word set comprises a plurality of range words;

obtaining a natural language-based data query request to be analyzed;

performing word segmentation on the data query request to obtain a first word set;

obtaining a word in the first word set that is consistent with the range word through filtering by using each of the keyword filters;

5 obtaining a query range of the data query request based on all words obtained through filtering;

filtering out the word in the first word set that is consistent with the range word to obtain a second word set;

performing semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result;

10 determining a data presentation dimension and a data presentation method corresponding to the data query request based on the semantic analysis result; and

outputting the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

10. The computer device according to claim 9, wherein the method further comprises:

15 obtaining historical data query intentions when each of the keyword filters fails to obtain the word in the first word set that is consistent with the range word through filtering;

determining a historical data query intention that has a highest degree of matching with the data presentation dimension and the data presentation method; and

20 obtaining a historical query range in the historical data query intention with the highest degree of matching as the query range of the data query request; wherein

a text font color representing the query range is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

11. The computer device according to claim 9, wherein the method further comprises:

25 obtaining historical data query intentions when the data presentation dimension corresponding to the data query request cannot be determined based on the semantic analysis result;

determining a historical data query intention that has a highest degree of matching with the query range and the data presentation method; and

obtaining a historical data presentation dimension in the historical data query intention with the highest degree of matching as the data presentation dimension of the data query request;
 5 wherein

a text font color representing the data presentation dimension is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

12. The computer device according to claim 9, wherein the method further comprises:

10 obtaining historical data query intentions when the data presentation method corresponding to the data query request cannot be determined based on the semantic analysis result;

determining a historical data query intention that has a highest degree of matching with the query range and the data presentation dimension; and

obtaining a historical data presentation method in the historical data query intention with the
 15 highest degree of matching as the data presentation method of the data query request; wherein

a text font color representing the data presentation method is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

13. The computer device according to claim 9, wherein the semantic analysis result
 20 comprises a plurality of word-semantics pairs, and the word-semantics pair comprises one word in the second word set and semantics of the word.

14. The computer device according to claim 13, wherein the method further comprises:

presetting semantics corresponding to each data presentation dimension in data content; and

the determining a data presentation dimension corresponding to the data query request based
 25 on the semantic analysis result comprises:

matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation dimension in the data content; and

when the semantics in the word-semantics pair are the same as semantics corresponding to a first data presentation dimension in the data content, using the first data presentation dimension as the data presentation dimension corresponding to the data query request.

15. A computer readable storage medium on which a computer program is stored, wherein
5 when the program is executed by a processor, the following steps of a natural-language data query intention determining method are implemented:

presetting a plurality of keyword filters representing a query range, wherein each of the keyword filters corresponds to a range word set, and the range word set comprises a plurality of range words;

10 obtaining a natural language-based data query request to be analyzed;

performing word segmentation on the data query request to obtain a first word set;

obtaining a word in the first word set that is consistent with the range word through filtering by using each of the keyword filters;

obtaining a query range of the data query request based on all words obtained through filtering;

15 filtering out the word in the first word set that is consistent with the range word to obtain a second word set;

performing semantic annotation on a word in the second word set based on a semantic knowledge base to generate a semantic analysis result;

20 determining a data presentation dimension and a data presentation method corresponding to the data query request based on the semantic analysis result; and

outputting the query range, the data presentation dimension, and the data presentation method as a standardized data query intention corresponding to the data query request.

16. The computer readable storage medium according to claim 15, wherein the method further comprises:

25 obtaining historical data query intentions when each of the keyword filters fails to obtain the word in the first word set that is consistent with the range word through filtering;

determining a historical data query intention that has a highest degree of matching with the data presentation dimension and the data presentation method; and

obtaining a historical query range in the historical data query intention with the highest degree of matching as the query range of the data query request; wherein

5 a text font color representing the query range is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

17. The computer readable storage medium according to claim 15, wherein the method further comprises:

10 obtaining historical data query intentions when the data presentation dimension corresponding to the data query request cannot be determined based on the semantic analysis result;

determining a historical data query intention that has a highest degree of matching with the query range and the data presentation method; and

15 obtaining a historical data presentation dimension in the historical data query intention with the highest degree of matching as the data presentation dimension of the data query request; wherein

a text font color representing the data presentation dimension is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

20 18. The computer readable storage medium according to claim 15, wherein the method further comprises:

obtaining historical data query intentions when the data presentation method corresponding to the data query request cannot be determined based on the semantic analysis result;

25 determining a historical data query intention that has a highest degree of matching with the query range and the data presentation dimension; and

obtaining a historical data presentation method in the historical data query intention with the highest degree of matching as the data presentation method of the data query request; wherein

a text font color representing the data presentation method is set to gray when the query range, the data presentation dimension, and the data presentation method are output as the standardized data query intention corresponding to the data query request.

19. The computer readable storage medium according to claim 15, wherein the semantic
5 analysis result comprises a plurality of word-semantics pairs, and the word-semantics pair comprises one word in the second word set and semantics of the word.

20. The computer readable storage medium according to claim 19, wherein the method further comprises:

presetting semantics corresponding to each data presentation dimension in data content; and
10 the determining a data presentation dimension corresponding to the data query request based on the semantic analysis result comprises:

matching the semantics in the word-semantics pair with the semantics corresponding to each data presentation dimension in the data content; and

when the semantics in the word-semantics pair are the same as semantics corresponding to a
15 first data presentation dimension in the data content, using the first data presentation dimension as the data presentation dimension corresponding to the data query request.