

SPECIFICATION

QUESTION DATA GENERATION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

The present application claims priority to the Chinese Patent Application No. 2018112046061, filed with the Chinese Patent Office on October 16, 2018, and entitled
5 "QUESTION DATA GENERATION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

10 The present application relates to a question data generation method and apparatus, a computer device, and a storage medium.

BACKGROUND

Questions are asked and answered manually in a conventional credit scenario. With the continuous development of computer technologies, systematic human-machine questioning and
15 answering gradually emerge. A terminal replaces a business person to question a user, and therefore costs of industry training for the business person decrease. However, with the systematic human-machine questioning and answering, question data is usually determined based on a pre-configured fixed question library template. In other words, in a credit interview process, a server generates a set of fixed question data based on the fixed question library
20 template, and displays the fixed question data in sequence by using a terminal.

However, the inventor realizes that, the question data in the credit interview process is fixed and not dynamic at present, disclosure of the credit interview question library may occur, and consequently the credit interview question library has low security.

SUMMARY

25 According to various embodiments disclosed in the present application, a question data generation method and apparatus, a computer device, and a storage medium are provided.

A question data generation method includes:

receiving user answer data that corresponds to current question data and that is sent by a terminal;

extracting a keyword from the user answer data based on a preset keyword extraction manner;

- 5 querying a built knowledge graph based on the keyword for a matched entity and attribute;
determining target question data based on the found entity and attribute; and
sending the target question data to the terminal for display.

A question data generation apparatus includes:

- 10 a receiving module, configured to receive user answer data that corresponds to current question data and that is sent by a terminal;

an extraction module, configured to extract a keyword from the user answer data based on a preset keyword extraction manner;

a query module, configured to query a built knowledge graph based on the keyword for a matched entity and attribute;

- 15 a generation module, configured to determine target question data based on the found entity and attribute; and

a sending module, configured to send the target question data to the terminal for display.

- 20 A computer device includes a memory and one or more processors, where the memory stores computer readable instructions, and when the computer readable instructions are executed by the one or more processors, the one or more processors are enabled to implement the steps of the question data generation method provided in any embodiment of the present application.

- 25 One or more non-volatile computer readable storage media storing computer readable instructions are provided, where when the computer readable instructions are executed by one or more processors, the one or more processors are enabled to implement the steps of the question data generation method provided in any embodiment of the present application.

Details of one or more embodiments of the present application are set forth in the following accompanying drawings and descriptions. Other features and advantages of the present application will become apparent from the specification, accompanying drawings, and claims.

BRIEF DESCRIPTION OF DRAWINGS

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

FIG. 1 is a diagram of an application scenario of a question data generation method according to one or more embodiments;

10 FIG. 2 is a schematic flowchart of a question data generation method according to one or more embodiments;

FIG. 3 is a schematic flowchart of a question data generation method in another embodiment;

15 FIG. 4 is a block diagram of a question data generation apparatus according to one or more embodiments;

FIG. 5 is a block diagram of a question data generation apparatus according to another embodiment; and

FIG. 6 is a block diagram of a computer device according to one or more embodiments.

20 DETAILED DESCRIPTION

To make the 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
25 embodiments described herein are merely used to illustrate the present application and are not intended to limit the present application.

A question data generation method provided in the present application may be applied to an application environment shown in FIG. 1. A terminal 102 communicates with a server 104 through a network. The server 104 receives user answer data fed back by the terminal 102 for current question data, extracts a keyword from the user answer data, queries a built knowledge
30 graph based on the extracted keyword for an entity and attribute that match the keyword, generates target question data based on the found entity and attribute, and sends the target

question data to the terminal 102 for display. The terminal 102 may be but is not limited to various personal computers, laptops, smartphones, tablets, and portable wearable devices. The server 104 may be implemented by an independent server or a server cluster composed of a plurality of servers.

5 In an embodiment, as shown in FIG. 2, a question data generation method is provided. An example in which the method is applied to the server in FIG. 1 is used for description. The method includes the following steps.

S202. Receive user answer data that corresponds to current question data and that is sent by a terminal.

10 The current question data is question data with which a user is currently asked, namely, question data corresponding to a current question-and-answer data processing process. The current question data may be specifically question data most recent from a current time, namely, latest question data. The question data may be specifically text data displayed by a display, or may be speech information displayed in a text-to-speech manner. The current question data
15 corresponds to the user answer text and user answering image that are currently fed back by the terminal. The user answer data is answer data fed back by the user for the current question data. The user answer data may be specifically answer data in a text form that is manually entered by the user into the terminal, or may be answer information in a speech form that is collected by the terminal when the user answers the current question data.

20 Specifically, when receiving the current question data sent by a server, the terminal displays the received current question data to the user in the text-to-speech manner or by using the display, and detects in real time the user answer data fed back by the user for the current question data. When detecting the user answer data fed back by the user for the current question data, the terminal sends the detected user answer data to the server.

25 In an embodiment, the server receives a question-and-answer instruction, obtains the corresponding current question data according to the received question-and-answer instruction, and sends the obtained current question data to the terminal. The server selects the current question data from a pre-stored preset question library according to the question-and-answer instruction. The server may alternatively obtain a corresponding question material from a
30 pre-stored question material set according to the question-and-answer instruction, and generate the corresponding current question data based on the obtained question material. The server may alternatively select an entity and a corresponding attribute from a built knowledge graph according to the question-and-answer instruction, and generate the current data based on the

selected entity and attribute. The preset question library is a question data set composed of a plurality of pieces of pre-generated question data. The question material set is a material set composed of a plurality of questions materials. For example, the question material is an identity card number, a native place, or a current residence.

5 In an embodiment, when displaying the current question data to the user, the terminal detects in real time an answering operation of the user for the current question data, and obtains the corresponding user answer data based on the detected answering operation. For example, the answering operation is an entry operation performed by the user on the terminal, or an oral operation of the user.

10 In an embodiment, when the received user answer data is answer data in the speech form, the server performs speech recognition on the user answer data, obtains answer speech text corresponding to the user answer data, and performs the following related steps in the question data generation method based on the extracted answer speech text. The speech recognition may be based on an existing speech recognition technology. Details are not described herein.

15 S204. Extract a keyword from the user answer data based on a preset keyword extraction manner.

The preset keyword extraction manner is a preset manner of extracting a keyword from the user answer data. The preset keyword extraction manner may be matching a keyword based on a preset keyword library; or performing word segmentation on the user answer data, and selecting
20 a keyword from a plurality of keywords obtained through word segmentation based on a preset selection condition; or extracting a keyword based on a trained keyword extraction model. The keyword extraction model may be specifically classified into a model obtained through training based on a supervised, semi-supervised, or unsupervised training manner.

Specifically, the server extracts the keyword from the received user answer data based on
25 the preset keyword extraction manner, to obtain the keyword corresponding to the user answer data. The server performs word segmentation on the user answer data, performs keyword matching between the keywords obtained through word segmentation and the preset keyword library, and uses a term successfully matching the preset keyword library as the keyword. The preset keyword library is a preset keyword set composed of a plurality of keywords.

30 In an embodiment, the server performs word segmentation on the user answer data to obtain a word segmentation result, calculates a feature weight for each term in the word segmentation result, and sorts terms in the word segmentation result based on feature weights, to further select the keyword based on a sorting result. Specifically, the server calculates a TF-IDF

weight for each term in the word segmentation result, sorts the terms in descending order of the TF-IDF weights, and selects a preset quantity of terms sorted ahead as keywords. A process of calculating a TF-IDF weight for each term includes the following: A term frequency of each term in the word segmentation result is first calculated, where the calculation may be performed with reference to the following formula:

Term frequency TF = Quantity of occurrence times of a term in a document/Total quantity of terms in the document;

Afterwards, an inverse document frequency IDF of each term is calculated, where the calculation may be performed with reference to the following formula:

$$\text{IDF} = \log \left(\frac{\text{Total quantity of documents in a corpus}}{(\text{Quantity of documents containing the term} + 1)} \right)$$

10 Inverse document frequency

Finally, the TF-IDF weight is obtained by calculating a product of the term frequency TF and the inverse document frequency IDF.

In an embodiment, the server inputs the user answer data into a pre-trained keyword extraction model for prediction, to obtain the keyword corresponding to the user answer data. The keyword extraction model is a model obtained through training based on a pre-obtained training sample set.

S206. Query a built knowledge graph based on the keyword for a matched entity and attribute.

The knowledge graph is a semantic network graph containing various entities or concepts and their relationships. Nodes in the knowledge graph are entity nodes corresponding to the entities. A line connected to any entity node represents an attribute of an entity corresponding to the any entity node. An entity corresponding to an entity node connected to the other end of the line connected to the any entity node is an attribute value corresponding to the any entity node. For example, if the entity corresponding to the any entity node is headquarters of XX Group and the line connected to the any entity node represents an address, the entity corresponding to the entity node connected to the other end of the line connected to the any entity node is Futian District, Shenzhen City. The entities in the knowledge graph may include a concept, a person name, a place name, an enterprise organization name, or the like. An attribute in the knowledge graph may be a relationship (entity relationship) between two entities connected by a line corresponding to the attribute, or an inherent feature of a corresponding entity.

It can be understood that the knowledge graph may be specifically a knowledge base containing a large quantity of 3-tuple. The 3-tuple includes an entity, an entity relationship, and an entity associated with the entity through the entity relationship. An entity relationship in each 3-tuple is used as an attribute between an entity and a corresponding associated entity, and a corresponding directed connection is established between the entity and the corresponding associated entity based on the attribute, to build a corresponding knowledge graph.

Specifically, the server matches the extracted keyword separately with each entity and attribute in the built knowledge graph, and determines an entity and attribute successfully matching the keyword as the entity and attribute that match the keyword. There are a plurality of keywords extracted from the user answer data. The server matches each extracted keyword with each entity in the knowledge graph, and determines a successfully matched entity as an entity found based on the keyword. Similarly, the server matches each extracted keyword with each attribute in the knowledge graph, and determines a successfully matched attribute as an attribute found based on the keyword.

In an embodiment, the server may match a synonym of the extracted keyword separately with each attribute in the knowledge graph, and determine a successfully matched attribute as an attribute matching the keyword. The server may alternatively match the found keyword with keywords of attributes in the knowledge graph, and determine a successfully matched attribute as an attribute matching the keyword.

In an embodiment, the server matches the extracted keyword with the entities in the knowledge graph, so that the keyword corresponds to a specific entity in the knowledge graph, and an entity matching the keyword is found in the knowledge graph. The server understands semantics expressed by the user answer data by using the NLP (natural language processing) technology, matches the understood semantics with the attributes in the knowledge graph to identify an attribute corresponding to the user answer data, and determines the attribute as an attribute matching the corresponding keyword. The server may alternatively match semantics expressed by the current question data with the attributes in the knowledge graph, to determine an attribute matching the extracted keyword.

In an embodiment, the keyword extracted from the user answer data can be understood as a name entity identified from the user answer data. The semantics expressed by the user answer data can be understood as an attribute corresponding to the name entity. The name entity includes a person name, a place name, an enterprise organization name, or the like. When a plurality of keywords are extracted, the server may determine the name entity based on the

plurality of keywords.

In an embodiment, the server queries the knowledge graph based on the identified name entity for a matched entity. The server calculates a match rate between the name entity and each entity in the knowledge graph, and determines an entity whose match rate reaches a preset threshold as the entity matching the name entity. The match rate is a match degree between the name entity and the corresponding entity. For example, when the name entity is an enterprise organization name, the entity matching the name entity may be the enterprise organization name, or may be another enterprise organization name corresponding to an enterprise organization corresponding to the enterprise organization name, such as an abbreviation or a full name. When the name entity is a place name, an entity matching the name entity may be the place name, or may be another place name corresponding to an address within a preset area from an address corresponding to the place name.

For example, assuming that the current question data is "where is your current residence?", the obtained user answer data is "I currently live in No. 58, Keyuan Road, Nanshan District." The name entity identified from the user answer data in the foregoing manner is "No. 58, Keyuan Road, Nanshan District". Based on the foregoing keyword extraction manner or semantics understanding manner, it can be determined that an attribute corresponding to the name entity is "address". The server queries the knowledge graph for an entity matching the name entity, and the entity may be "No. 58, Keyuan Road, Nanshan District", "No. 59, Keyuan Road, Nanshan District", or the like.

In an embodiment, when the current question data is a yes-or-no question, the server may determine content text of the user answer data as the keyword corresponding to the user answer data. Further, the server may determine the semantics represented by the user answer data based on the current question data and the keyword extracted from the user answer data, and query the built knowledge graph based on the semantics for an entity and attribute matching the semantics. For example, assuming that the current question data is "there is an XX Bank at No. 59, Keyuan Road, Nanshan District, yes or no?", and the corresponding user answer data is "yes", the entity and attribute determined based on the semantics represented by the user answer data may respectively be "No. 59, Keyuan Road, Nanshan District" and "address".

S208. Determine target question data based on the found entity and attribute.

The target question data is question data used to ask the user following the current question data. The target question data may be specifically text data that can be displayed by a display, or may be speech information that can be displayed in a text-to-speech manner.

Specifically, based on the found entity and attribute, the server queries the corresponding knowledge graph for entities associated with the entity through the attribute. The server generates the corresponding target question data based on the found entity and attribute and a corresponding associated entity.

5 For example, based on the foregoing example, assuming that "No. 59, Keyuan Road, Nanshan District" is the entity found in the knowledge graph based on the identified name entity, and "XX Bank" is the associated entity found in the knowledge graph based on the attribute "address" corresponding to the name entity, the generated target question data may be "there is an XX Bank at No. 59, Keyuan Road, Nanshan District, yes or no?" or "there is an XX Bank
10 near No. 58, Keyuan Road, Nanshan District, yes or no?"

In an embodiment, the server queries the built knowledge graph based on the keyword extracted from the user answer data for an entity matching the keyword, and based on the knowledge graph, determines one or more lines connected to the entity. The one or more lines represent an attribute or attributes of the entity. The server may select an attribute from the one
15 or more attributes, and generate the target question data based on the selected attribute and the corresponding entity. In an embodiment, the server queries the knowledge graph for one or more entities associated with the entity through the one or more lines. The one or more associated entities are an attribute value or attribute values corresponding to the attribute or attributes. The server selects an attribute from the one or more attributes, selects an associated entity
20 corresponding to the selected attribute from the corresponding one or more associated entities, and generates the target question data based on the selected attribute, the associated entity, and the corresponding entity.

For example, assuming that the current question data is "which company do you work for?", the obtained user answer data is "I work for XX Group." The entity found based on the keyword
25 in the user answer data in the foregoing manner is "XX Group", and the attribute found in the knowledge graph based on the entity may be "address", "legal person", "establishment time", or the like. The target question data generated in the foregoing manner may be "who is the legal person of XX Group?", "the legal person of XX Group is Zhang San, yes or no?", or the like.

S210. Send the target question data to the terminal for display.

30 Specifically, the server sends the target question data determined based on the found entity and attribute to the terminal through a network. The terminal displays the received target question data to the corresponding user by using the display or in the text-to-speech manner.

In an embodiment, when displaying the received target question data to the user in the

text-to-speech manner or by using the display, the terminal detects in real time the user answer data fed back by the user for the target question data, and sends the detected user answer data to the server, so that the server continues to perform related steps of the question data generation method based on the received user answer data.

5 In an embodiment, the server sends the target question data to a user terminal, to display the target question data to the corresponding user through the user terminal. The server may alternatively send the target question data to a business person's terminal for display.

In the foregoing question data generation method, the user answer data fed back by the terminal for the current question data is received; the corresponding keyword is extracted from the user answer data based on the preset keyword extraction manner; the built knowledge graph is queried based on the extracted keyword for the entity and attribute matching the keyword; and the corresponding target question data is further generated based on the found entity and attribute. In this way, accuracy of the target question data is improved. Target question data of a next question is dynamically determined based on the user answer data and the built knowledge graph, to ensure uncertainty of the target question data and effectively avoid disclosure of a question library composed of various question data, so that security of the question library is improved.

In an embodiment, step S208 includes: based on the found entity and attribute, querying the knowledge graph for a plurality of entities associated with the entity through the attribute; selecting a target associated entity from the associated entities; and generating the target question data based on the entity, the attribute, and the target associated entity.

An associated entity is an entity associated with another entity through a line (attribute) in a knowledge graph. It can be understood that both the entity and the associated entity are entity nodes in the knowledge graph. The entity and the associated entity are associated with each other through the attribute, or to be specific, the attribute is used to represent an entity relationship between the entity and the associated entity. The two entities associated with each other through the attribute may be mutually associated entities. The associated entity may be an attribute value corresponding to the entity, or may be another entity associated with the entity through the entity relationship represented through the attribute.

30 Specifically, based on the found entity and attribute, the server queries the knowledge graph for a plurality of entities associated with the entity through the attribute. The server selects the target associated entity from the plurality of found associated entities, and generates the target question data based on the target associated entity and the corresponding entity and

attribute.

In an embodiment, the server may randomly select an associated entity from the plurality of found associated entities as the target associated entity. The server may alternatively obtain a user identifier corresponding to the received user answer data, calculate a total quantity of the found associated entities, calculate a hash value based on the calculated total quantity for the user identifier, and further select the target associated entity from the plurality of associated entities based on the calculated hash value. The server may alternatively select an associated entity from the plurality of found associated entities as the target associated entity in sequence in a polling manner.

In the foregoing embodiment, the target associated entity is selected from the plurality of found associated entities, and the target question data is further generated based on the target associated entity and the corresponding entity and attribute, thereby further increasing the uncertainty of the target question data, and further improving the security of the question library.

In an embodiment, step S204 includes: preprocessing the user answer data; and extracting the keyword from the preprocessed user answer data in a preset keyword library-based matching manner.

The preprocessing includes word segmentation and stop word removal. Word segmentation is a process of segmenting the user answer data in the text form into individual terms. There are various word segmentation algorithms corresponding to word segmentation, for example, a word segmentation algorithm based on character string matching, a word segmentation algorithm based on semantic analysis, a word segmentation algorithm based on statistics, or the like. For example, the word segmentation algorithm based on character string matching is positive maximum matching algorithm, inverse maximum matching algorithm, minimum segmentation algorithm, or bidirectional maximum matching algorithm. A trained word segmentation model may alternatively be used for performing word segmentation on the user answer data. The word segmentation model may be a hidden Markov model, a CRF (conditional random field algorithm) model, or the like. The stop word removal means that in information retrieval, to save storage space and improve retrieval efficiency, some letters or words are automatically filtered out before or after processing of natural language data (or text), for example, modal particles, polite words, prepositions, or conjunctions. The preset keyword library is a preset keyword set composed of a plurality of preset keywords. The preset keyword library may also include a synonym of each preset keyword.

Specifically, the server performs word segmentation on the received user answer data to

obtain the word segmentation result, and performs stop word removal on the word segmentation result to obtain a corresponding candidate keyword set. The server may match each term in the word segmentation result with a pre-built stop word library, determine a term successfully matched the stop word library as a stop word, and remove the stop word from the word segmentation result. The server uses terms in the word segmentation result obtained after the stop word removal as candidate keywords, to obtain the corresponding candidate keyword set. Further, the server matches each candidate keyword in the candidate keyword set with a preset keyword in the preset keyword library and a synonym of the preset keyword. When the match is successful, it indicates that the successfully matched candidate keyword is a preset keyword in the preset keyword library or a synonym of the preset keyword. The server determines the candidate keyword as the keyword extracted from the preprocessed user answer data.

In an embodiment, the server uses the NLP (natural language processing) technology to process the user answer data, and performs semantic analysis and understanding on the processed user answer data, to obtain the corresponding keyword. In an embodiment, the server preprocesses the user answer data, and extracts the keyword from the preprocessed user answer data in a matching manner that is based on a dictionary and a regular expression. Specifically, the server matches each candidate keyword obtained through the preprocessing with a preset keyword in a preset dictionary, and determines a successfully matched candidate keyword as the extracted keyword.

In the foregoing embodiment, the keyword is extracted based on the preset keyword library, so that the keyword extracted from the user answer data is more accurate.

In an embodiment, before step S206, the question data generation method further includes: obtaining target data; identify target entities in the target data and entity relationships between the target entities; and building the knowledge graph in a preset building manner based on the target entities and the corresponding entity relationships.

The target data is raw data used to build the knowledge graph. The target data may be specifically raw data in one or more designated fields. For example, the designated fields are the financial field and the manufacturing field. Raw data in the financial field is used as an example. The financial field may include industries such as banking, securities, insurance, trusts, credit, and funds. Each industry may include a plurality of enterprise organizations. Each enterprise organization has corresponding enterprise raw data. The enterprise raw data may include internal data, external data, and the like. The internal data is data stored in an enterprise local database, for example, operation data and service data. The external data is data that may be

crawled from other storage space by using a third-party platform, for example, industrial and commercial data, banking regulatory data, data of the People's Bank of China, and a company annual report. Enterprise raw data of a group enterprise may further include raw data of each subsidiary, for example, an investment relationship, establishment time, and an address.

5 Enterprise raw data including bank data may further include an address or the like of each branch.

Specifically, when receiving a knowledge graph building instruction, the server obtains disclosed raw data from various channels according to the received building instruction, and obtains pre-stored and undisclosed raw data from a local database, to obtain the corresponding

10 target data. The server processes the obtained target data, to further identify the target entities in the target data and the entity relationships between the target entities. The server uses an entity relationship between any two target entities as an attribute between the two target entities, and establishes a corresponding directed connection between the two entities based on the attribute. The server performs the foregoing steps for each target entity and a corresponding entity

15 relationship that are identified from the target data, to build the knowledge graph corresponding to the target data.

In an embodiment, the server may use a data warehouse technology to process the obtained target data. The server saves relational data in the knowledge graph built in the foregoing manner to a graph database. The graph database may be a Neo4j database (a high-performance

20 NoSQL graph database).

In an embodiment, the target data includes raw data in a plurality of designated fields. For raw data in each designated field, the server may use the field as a vertex (core node) of the knowledge graph, and determine, based on the entity relationships identified from the target data, associated nodes connected to the core node. Similarly, the server determines an associated node

25 connected to each node based on the entity relationships, uses the entity relationships between the nodes as attributes, and builds the knowledge graph based on the determined nodes and the corresponding attributes. Both the core node and the associated nodes are nodes (entity nodes) in the knowledge graph. In an embodiment, the server builds a knowledge graph corresponding to each designated field in the foregoing manner, and combines knowledge graphs

30 corresponding to any two designated fields based on an intersecting entity or attribute between the two designated fields, to obtain a cross-field knowledge graph. The intersecting entity or attribute between the two designated fields is an entity or attribute contained in both the knowledge graphs corresponding to the two designated fields.

In the foregoing embodiments, the corresponding entities and entity relationships are identified based on the target data. The corresponding knowledge graph is further built based on the identified entities and entity relationships. Therefore, the built knowledge graph can be directly used to dynamically generate the question data, and question data generation efficiency is improved.

In an embodiment, the user answer data includes answer speech information. Step S204 includes: performing speech recognition on the answer speech information to obtain corresponding answer speech text; and extracting the keyword from the answer speech text based on the preset keyword extraction manner.

The answer speech information is speech information collected when the user performs an answering operation for the current question data. The answer speech text is speech text content carried in the answer speech information.

Specifically, the server obtains the corresponding answer speech information from the received user answer data, and performs speech recognition on the obtained answer speech information, to extract the corresponding answer speech text from the answer speech information. The server performs word segmentation on the extracted answer speech text, performs keyword matching between terms obtained through word segmentation and the preset keyword library, and uses a term successfully matched the preset keyword library as the keyword extracted from the answer speech text.

In an embodiment, the server extracts the corresponding answer speech text from the answer speech information by using a pre-trained speech recognition model, and extracts the corresponding keyword from the answer speech text by using the pre-trained keyword extraction model. It can be understood that the server may alternatively extract the corresponding answer speech text from the answer speech information based on another speech recognition technology in the prior art. Similarly, the server may alternatively extract the keyword based on another keyword extraction technology in the prior art. Details are not described herein.

In an embodiment, the server performs voiceprint recognition on the answer speech information to extract a corresponding target voiceprint feature, and performs identity authentication based on the target voiceprint feature on the user performing the answering operation for the current question data. When the identity authentication succeeds, the server continues to perform speech recognition on the answer speech information to obtain the corresponding answer speech text.

In the foregoing embodiment, when the user answer data includes the answer speech

information, the corresponding answer speech text is identified based on the speech recognition technology. The keyword is further extracted from the answer speech text. In this way, user answer data processing efficiency is improved, and the question data generation efficiency is improved.

5 In an embodiment, the user answer data includes user answer text and a user answering image. Before step S206, the question data generation method further includes: determining an answer score corresponding to the user answer text, and determining an expression score corresponding to the user answering image; determining a corresponding comprehensive score based on the answer score and the expression score; and querying the built knowledge graph
10 based on the keyword for the matched entity and attribute when the comprehensive score is less than a preset score threshold.

 The user answer text is answer data in the text form. The terminal may obtain the corresponding user answer text based on a manual entry operation of the user, and may alternatively use speech content as the user answer text, where the speech content is identified
15 from speech information collected when the user performs the answering operation. The user answering image is image information collected when the user feeds back the answer data for the current question data. The image information may be specifically information collected by an image collector in a form of a picture, a video, or the like. The user answering image may include but is not limited to a user answering picture and a user answering video. For example,
20 the user answering picture is a picture collected when the user is answering a question, including a user facial picture. For example, the user answering video is a video collected when the user is answering a question, to be specific, a video collected when the user is performing an answering operation. The image collector may be a camera. The camera may be a camera equipped on the terminal, or may be an independent component connected to the terminal point to point. The
25 user answering image corresponds to the user answer text.

 Specifically, the server obtains the corresponding user answer text and the user answering image from the received user answer data. The server determines the answer score corresponding to the user answer text in a preset answer score determination manner, and determines the expression score corresponding to the user answering image in a preset
30 expression score determination manner. The server determines the comprehensive score corresponding to the user answer data based on the determined answer score and expression score, and compares the determined comprehensive score with the preset score threshold. Based on the keyword extracted from the user answer text contained in the user answer data, the server queries the built knowledge graph for the matched entity and attribute when the comprehensive

score is less than the preset score threshold.

The preset answer score determining manner may be determining the answer score based on a match rate between the user answer text and a corresponding preset answer, and a preset answer score corresponding to the current question data; or determining the corresponding answer score by matching a keyword in the user answer text with a preset keyword; or obtaining the corresponding answer score by entering the user answer text into a trained answer score prediction model. The preset expression score determining manner may be extracting a corresponding user micro-expression from the user answering image, and determining the expression score corresponding to the user answering image based on the extracted user micro-expression and a micro-expression recognition technology.

In an embodiment, the server obtains a plurality of frames of user facial pictures based on the user answering image, determines expression scores respectively corresponding to the plurality of user facial pictures based on the micro-expression recognition technology, and determines the expression score corresponding to the user answering image based on the plurality of expression scores. In an embodiment, when the expression score corresponding to the user answering image is less than a preset expression score threshold, the server queries the built knowledge graph based on the corresponding keyword for the matched entity and attribute.

In the foregoing embodiment, when the comprehensive score is less than the preset score threshold, it indicates that correctness of the user answer data is doubtful. In this case, an entity and attribute related to the current question data are selected based on the built knowledge graph, and the corresponding target question data is dynamically generated, so that the dynamically generated target question data is more accurate.

In an embodiment, the user answer data corresponds to a user identifier. The question data generation method further includes: calculating a total amount of question data corresponding to the user identifier; stopping a question data generation process when the calculated total amount reaches a preset total amount; determining a total score corresponding to the user identifier based on comprehensive scores corresponding to the question data corresponding to the user identifier; and pushing the total score to the terminal for display.

Specifically, the server receives the user answer data fed back by the terminal for the user identifier and the current question data, calculates the total amount of the question data corresponding to the user identifier, and compares the calculated total amount with the preset total amount. When the calculated total amount reaches the preset total amount, the server stops the current question data generation process. The server determines a comprehensive score

corresponding to the current question data in the foregoing manner, obtains the comprehensive scores corresponding to the existing the question data corresponding to the user identifier, and performs direct summation or weighted summation on the currently determined comprehensive score and the plurality of obtained comprehensive scores, to determine the total score
 5 corresponding to the user identifier, and send the total score to the terminal for display.

In an embodiment, the server generates corresponding prompt information based on the calculated total score and a preset total score threshold, and pushes the prompt information to the terminal. For example, when the foregoing method is applied to a question-and-answer session in a credit interview process, the server may generate prompt information indicating that
 10 the credit interview succeeds or fails. The interview means that a user identity is reviewed in a credit service handling process. When the foregoing method is applied to a job interview session, the server may generate prompt information indicating that the interview succeeds or fails.

In an embodiment, the server determines the comprehensive score corresponding to the current question data in the foregoing manner, and updates an existing total score corresponding
 15 to the user identifier based on the comprehensive score. When an updated total score reaches the preset total score threshold, the server stops the current question data generation process.

As shown in FIG. 3, in an embodiment, a question data generation method is provided, and the method specifically includes the following steps:

S302. Receive user answer data that corresponds to current question data and that is sent by
 20 a terminal.

S304. Preprocess the user answer data.

S306. Extract a keyword from the preprocessed user answer data in a preset keyword library-based matching manner.

S308. Obtain target data.

25 S310. Identify target entities in the target data and entity relationships between the target entities.

S312. Build a knowledge graph in a preset building manner based on the target entities and the corresponding entity relationships.

S314. Query a built knowledge graph based on the keyword for a matched entity and
 30 attribute.

S316. Based on the found entity and attribute, query the knowledge graph for a plurality of

entities associated with the entity through the attribute.

S318. Select a target associated entity from the associated entities.

S320. Generate target question data based on the entity, the attribute, and the target associated entity.

5 S322. Send the target question data to the terminal for display.

In the foregoing embodiment, the keyword is extracted from the user answer data corresponding to the current question data. The matched entity and attribute are queried in the built knowledge graph based on the keyword, and the associated entities are determined. The target question data is further dynamically generated based on the entity, the attribute, and the
10 corresponding associated entity, and sent to the terminal for display. In this way, uncertainty of the target question data is increased, and security of a question library is improved.

In an embodiment, the foregoing question data generation method may be applied to a question-and-answer session of a credit interview, or may be applied to an interview part, or may be applied to a question-and-answer session of inspection and interrogation of public
15 security or customs, or the like.

It should be understood that although the steps in the flowcharts of FIG. 2 and FIG. 3 are shown in sequence indicated by arrows, these steps are not necessarily performed in such sequences. Unless expressly stated herein, these steps are not limited to a strict execution sequence, and may be performed in another sequence. In addition, at least some of the steps in
20 FIG. 2 and FIG. 3 may include a plurality of substeps or stages. These substeps or stages are not necessarily performed and completed at the same moment, but may be performed at different moments. These substeps or stages are not necessarily performed in sequence, either, but may be performed in turn or alternately with other steps or at least some substeps or stages of the other steps.

25 In an embodiment, as shown in FIG. 4, a question data generation apparatus 400 is provided, including a receiving module 402, an extraction module 404, a query module 406, a generation module 408, and a sending module 410.

The receiving module 402 is configured to receive user answer data that corresponds to current question data and that is sent by a terminal.

30 The extraction module 404 is configured to extract a keyword from the user answer data based on a preset keyword extraction manner.

The query module 406 is configured to query a built knowledge graph based on the keyword for a matched entity and attribute.

The generation module 408 is configured to determine target question data based on the found entity and attribute.

5 The sending module 410 is configured to send the target question data to the terminal for display.

10 In an embodiment, the generation module 408 is further configured to: based on the found entity and attribute, query the knowledge graph for a plurality of entities associated with the entity through the attribute; select a target associated entity from the associated entities; and generate the target question data based on the entity, the attribute, and the target associated entity.

In an embodiment, the extraction module 404 is further configured to: preprocess the user answer data; and extract the keyword from the preprocessed user answer data in a preset keyword library-based matching manner.

15 As shown in FIG. 5, in an embodiment, the foregoing question data generation apparatus 400 further includes a knowledge graph building module 412.

20 The knowledge graph building module 412 is configured to: obtain target data; identify target entities in the target data and entity relationships between the target entities; and build the knowledge graph in a preset building manner based on the target entities and the corresponding entity relationships.

In an embodiment, the user answer data includes answer speech information. The extraction module 404 is further configured to: perform speech recognition on the answer speech information to obtain corresponding answer speech text; and extract the keyword from the answer speech text based on the preset keyword extraction manner.

25 In an embodiment, the user answer data includes user answer text and a user answering image. The foregoing question data generation apparatus 400 further includes a determining module 414.

30 The determining module 414 is configured to: determine an answer score corresponding to the user answer text, and determine an expression score corresponding to the user answering image; determine a corresponding comprehensive score based on the answer score and the expression score; and enable the query module 406 to query the built knowledge graph based on

the keyword for the matched entity and attribute when the comprehensive score is less than a preset score threshold.

In an embodiment, the user answer data corresponds to a user identifier. The determining module 414 is further configured to: calculate a total amount of question data corresponding to the user identifier; stop a question data generation process when the calculated total amount reaches a preset total amount; determine a total score corresponding to the user identifier based on comprehensive scores corresponding to the question data corresponding to the user identifier; and push the total score to the terminal for display.

For a specific limitation on the question data generation apparatus, refer to the limitation on the question data generation method above. Details are not described herein. All or some of the modules in the foregoing question data generation apparatus may be implemented by using software, hardware, or a combination thereof. The foregoing modules may be built in or independent of a processor of a computer device in a hardware form, or may be stored in a memory of a computer device in a software form, so that the processor invokes and performs operations corresponding to the foregoing modules.

In an embodiment, a computer device is provided. The computer device may be a server. An internal structural diagram of the computer device may be shown in FIG. 6. The computer device includes a processor, a memory, a network interface, and a database that are connected through a system bus. The processor of the computer device is configured to provide computing and control capabilities. The memory of the computer device includes a non-volatile storage medium and an internal memory. The non-volatile storage medium stores an operating system, computer readable instructions, and a database. The internal memory provides an environment for running the operating system and the computer readable instructions in the non-volatile storage medium. The database of the computer device is configured to store a built knowledge graph. The network interface of the computer device is configured to communicate with an external terminal through a network connection. The computer readable instructions are executed by the processor to implement a question data generation method.

A person skilled in the art may understand that the structure shown in FIG. 6 is merely a block diagram of a partial structure related to the solution of the present application, and does not constitute a limitation on the computer device to which the solution of the present application is applied. Specifically, the computer device may include more or fewer components than those shown in the figure, or combine some components, or have different component arrangements.

A computer device includes a memory and one or more processors, where the memory stores computer readable instructions, and when the computer readable instructions are executed by the one or more processors, the one or more processors are enabled to implement the steps of the question data generation method provided in any embodiment of the present application.

5 One or more non-volatile computer readable storage media storing computer readable instructions are provided, where when the computer readable instructions are executed by one or more processors, the one or more processors are enabled to implement the steps of the question data generation method provided in any embodiment of the present application.

10 A person of ordinary skill in the art can understand that all or some of the processes of the methods in the embodiments may be implemented by the computer readable instructions instructing related hardware. The computer readable instructions may be stored in a non-volatile computer readable storage medium. When the computer readable instructions are executed, the processes of the method embodiments are performed. Any reference to a memory, a storage, a database, or another medium used in the embodiments provided in the present application may
15 include a non-volatile memory and/or volatile memory. The non-volatile memory may include a read-only memory (ROM), a programmable ROM (PROM), an electrically programmable ROM (EPROM), an electrically erasable programmable ROM (EEPROM), or a flash memory. The volatile memory may include a random access memory (RAM) or an external cache memory. By way of illustration but not limitation, the RAM is available in various forms, such as a static
20 RAM (SRAM), a dynamic RAM (DRAM), a synchronous DRAM (SDRAM), a double data rate SDRAM (DDR SDRAM), an enhanced SDRAM (ESDRAM), a synchlink (Synchlink) DRAM (SLDRAM), a Rambus (Rambus) direct RAM (RDRAM), a direct Rambus dynamic RAM (DRDRAM), and a Rambus dynamic RAM (RDRAM).

25 The technical features of the foregoing embodiments may be arbitrarily combined. For brevity of description, not all possible combinations of the technical features of the foregoing embodiments are described. However, it should be considered that these combinations of the technical features all fall within the scope of this specification provided that these combinations are not contradictory.

30 The foregoing embodiments merely describe several implementations of the present application, and description of the implementations is relatively specific and detailed, but shall not be understood as a limitation on the scope of the present invention. It should be noted that a person of ordinary skill in the art may make several variations and improvements without departing from the idea of the present application, and the variations and improvements shall

fall within the protection scope of the present application. Therefore, the protection scope of the present application shall be subject to the protection scope of the claims.

CLAIMS

What is claimed is:

1. A question data generation method, comprising:

receiving user answer data that corresponds to current question data and that is sent by a terminal;

5 extracting a keyword from the user answer data based on a preset keyword extraction manner;

querying a built knowledge graph based on the keyword for a matched entity and attribute;

determining target question data based on the found entity and attribute; and

sending the target question data to the terminal for display.

10 2. The method according to claim 1, wherein the determining target question data based on the found entity and attribute comprises:

based on the found entity and attribute, querying the knowledge graph for a plurality of entities associated with the entity through the attribute;

selecting a target associated entity from the associated entities; and

15 generating the target question data based on the entity, the attribute, and the target associated entity.

3. The method according to claim 1, wherein the extracting a keyword from the user answer data based on a preset keyword extraction manner comprises:

preprocessing the user answer data; and

20 extracting the keyword from the preprocessed user answer data in a preset keyword library-based matching manner.

4. The method according to claim 1, wherein before the querying a built knowledge graph based on the keyword for a matched entity and attribute, the method further comprises:

obtaining target data;

25 identifying target entities in the target data and entity relationships between the target entities; and

building the knowledge graph in a preset building manner based on the target entities and the corresponding entity relationships.

5 5. The method according to any one of claims 1 to 4, wherein the user answer data comprises answer speech information; and the extracting a keyword from the user answer data based on a preset keyword extraction manner comprises:

performing speech recognition on the answer speech information to obtain corresponding answer speech text; and

extracting the keyword from the answer speech text based on the preset keyword extraction manner.

10 6. The method according to any one of claims 1 to 4, wherein the user answer data comprises user answer text and a user answering image; and before the querying a built knowledge graph based on the keyword for a matched entity and attribute, the method further comprises:

15 determining an answer score corresponding to the user answer text, and determining an expression score corresponding to the user answering image;

determining a corresponding comprehensive score based on the answer score and the expression score; and

querying the built knowledge graph based on the keyword for the matched entity and attribute when the comprehensive score is less than a preset score threshold.

20 7. The method according to claim 6, wherein the user answer data corresponds to a user identifier; and the method further comprises:

calculating a total amount of question data corresponding to the user identifier;

stopping a question data generation process when the calculated total amount reaches a preset total amount;

25 determining a total score corresponding to the user identifier based on comprehensive scores corresponding to the question data corresponding to the user identifier; and

pushing the total score to the terminal for display.

8. A question data generation apparatus, comprising:

a receiving module, configured to receive user answer data that corresponds to current

question data and that is sent by a terminal;

an extraction module, configured to extract a keyword from the user answer data based on a preset keyword extraction manner;

5 a query module, configured to query a built knowledge graph based on the keyword for a matched entity and attribute;

a generation module, configured to determine target question data based on the found entity and attribute; and

a sending module, configured to send the target question data to the terminal for display.

10 9. A computer device, comprising a memory and one or more processors, wherein the memory stores computer readable instructions, and when the computer readable instructions are executed by the one or more processors, the one or more processors are enabled to perform the following steps:

receiving user answer data that corresponds to current question data and that is sent by a terminal;

15 extracting a keyword from the user answer data based on a preset keyword extraction manner;

querying a built knowledge graph based on the keyword for a matched entity and attribute;

determining target question data based on the found entity and attribute; and

sending the target question data to the terminal for display.

20 10. The computer device according to claim 9, wherein the determining target question data based on the found entity and attribute comprises:

based on the found entity and attribute, querying the knowledge graph for a plurality of entities associated with the entity through the attribute;

selecting a target associated entity from the associated entities; and

25 generating the target question data based on the entity, the attribute, and the target associated entity.

11. The computer device according to claim 9, wherein the extracting a keyword from the user answer data based on a preset keyword extraction manner comprises:

preprocessing the user answer data; and

extracting the keyword from the preprocessed user answer data in a preset keyword library-based matching manner.

12. The computer device according to claim 9, wherein when the computer readable instructions are executed by the processor, the processor further performs the following steps before querying the built knowledge graph based on the keyword for the matched entity and attribute:

obtaining target data;

identifying target entities in the target data and entity relationships between the target entities; and

building the knowledge graph in a preset building manner based on the target entities and the corresponding entity relationships.

13. The computer device according to any one of claims 9 to 12, wherein the user answer data comprises answer speech information; and the extracting a keyword from the user answer data based on a preset keyword extraction manner comprises:

performing speech recognition on the answer speech information to obtain corresponding answer speech text; and

extracting the keyword from the answer speech text based on the preset keyword extraction manner.

14. The computer device according to any one of claims 9 to 12, wherein the user answer data comprises user answer text and a user answering image; and when the computer readable instructions are executed by the processor, the processor further performs the following steps before querying the built knowledge graph based on the keyword for the matched entity and attribute:

determining an answer score corresponding to the user answer text, and determining an expression score corresponding to the user answering image;

determining a corresponding comprehensive score based on the answer score and the expression score; and

querying the built knowledge graph based on the keyword for the matched entity and attribute when the comprehensive score is less than a preset score threshold.

15. The computer device according to claim 14, wherein the user answer data corresponds

to a user identifier; and when executing the computer readable instructions, the processor further performs the following steps:

calculating a total amount of question data corresponding to the user identifier;

5 stopping a question data generation process when the calculated total amount reaches a preset total amount;

determining a total score corresponding to the user identifier based on comprehensive scores corresponding to the question data corresponding to the user identifier; and

pushing the total score to the terminal for display.

10 16. One or more non-volatile computer readable storage media storing computer readable instructions, wherein when the computer readable instructions are executed by one or more processors, the one or more processors are enabled to perform the following steps:

receiving user answer data that corresponds to current question data and that is sent by a terminal;

15 extracting a keyword from the user answer data based on a preset keyword extraction manner;

querying a built knowledge graph based on the keyword for a matched entity and attribute;

determining target question data based on the found entity and attribute; and

sending the target question data to the terminal for display.

20 17. The one or more storage media according to claim 16, wherein the determining target question data based on the found entity and attribute comprises:

based on the found entity and attribute, querying the knowledge graph for a plurality of entities associated with the entity through the attribute;

selecting a target associated entity from the associated entities; and

25 generating the target question data based on the entity, the attribute, and the target associated entity.

18. The one or more storage media according to claim 16, wherein the extracting a keyword from the user answer data based on a preset keyword extraction manner comprises:

preprocessing the user answer data; and

extracting the keyword from the preprocessed user answer data in a preset keyword library-based matching manner.

19. The one or more storage media according to claim 16, wherein when the computer readable instructions are executed by the one or more processors, the one or more processors
5 further perform the following steps before querying the built knowledge graph based on the keyword for the matched entity and attribute:

obtaining target data;

identifying target entities in the target data and entity relationships between the target entities; and

10 building the knowledge graph in a preset building manner based on the target entities and the corresponding entity relationships.

20. The one or more storage media according to any one of claims 16 to 19, wherein the user answer data comprises answer speech information; and the extracting a keyword from the user answer data based on a preset keyword extraction manner comprises:

15 performing speech recognition on the answer speech information to obtain corresponding answer speech text; and

extracting the keyword from the answer speech text based on the preset keyword extraction manner.