

MEDICAL KNOWLEDGE GRAPH QUESTION ANSWERING PROCESSING METHOD, DEVICE, APPARATUS, AND STORAGE MEDIUM

This application claims priority to Chinese Patent Application No. 2019101655569.4, filed on July 19, 2019 and entitled “Medical Knowledge Graph Question Answering Processing Method, Device, Apparatus, and Storage Medium”, the entire content of which is incorporated herein by reference.

TECHNICAL FIELD

This application relates to artificial intelligence natural language processing technologies, and more particularly, to a medical knowledge graph question answering processing method, device, apparatus, and storage medium.

BACKGROUND

Knowledge graph is also called scientific knowledge graph. In the library and information industry, the knowledge graph is called knowledge domain visualization or knowledge domain mapping map, which is a series of different graphs showing the relationship between knowledge development process and structure. Because the knowledge graph can provide high-quality structured data, knowledge graphs and question answering systems based on knowledge graphs, such as automatic question answering, search engines, and information extraction are used in more and more fields. A typical knowledge graph is usually expressed in the triple form including a head entity, a relationship, and a tail entity (for example,

Yao Ming, nationality, China). The expression of this instance reflects the fact that Yao Ming's nationality is Chinese.

The inventor realizes that in the medical field, constructing an automatic knowledge graph question answering processing system can effectively help patients or healthy users develop autonomy and prevention, and can alleviate the problem of hospitals' shortage of public medical resources. In practical applications, due to the particularity of the medical field, the requirement for the accuracy of the knowledge question answering system is relatively high. However, the current knowledge graph question answering technology is still in the stage of exploration and development, and most of the results and progress are mainly based on academic papers. The specific plan is: according to the questions raised by the user, the corresponding paper or website can be obtained through keyword search in the database; the user clicks on the specific content of the paper to find the content he needs, which results in a poor efficiency of processing the problem and thus cannot meet the user's requirements.

Therefore, it is an urgent technical problem to provide an effective knowledge graph question answering processing solution in the medical field.

SUMMARY OF THE DISCLOSURE

Therefore, the present disclosure provides a medical knowledge graph question answering processing method, device, apparatus, and storage medium, which are capable of providing the efficiency of processing questions from users and thus meeting the users' requirements.

The present disclosure provides a medical knowledge graph question answering processing method, including:

obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

obtaining a start position and an end position of each medical entity in the sentence to be processed;

according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model; and

according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

The present disclosure further provides an electronic device, including:

a recognition module configured to obtain a sentence to be processed and to recognize a medical entity in the sentence to be processed;

an obtaining module configured to obtain a start position and an end position of each medical entity in the sentence to be processed;

a determining module configured to, according to the medical entity and the corresponding start position and end position, compare the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

a processing module configured to perform relationship analysis on the sentence to be processed, and obtain a relationship corresponding to the

sentence to be processed based on a relationship matching model; and

an output module configured to determine an answer corresponding to the sentence to be processed according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, and output the answer.

The present disclosure further provides an apparatus including a storage, a processor, and a computer program stored in the storage and capable of running on the processor, wherein when the processor implements the computer program, a medical knowledge graph question answering processing method is performed, and the method includes the following steps:

obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

obtaining a start position and an end position of each medical entity in the sentence to be processed;

according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model; and

according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

The present disclosure further provides a computer-readable storage

medium on which a computer program is stored, wherein when the computer program is executed by a processor, a medical knowledge graph question answering processing method is implemented, including the following steps:

- obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

- obtaining a start position and an end position of each medical entity in the sentence to be processed;

- according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

- performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model; and

- according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

The medical knowledge graph question answering processing method, device, apparatus, and storage medium provided in the present disclosure recognize the medical entity in the sentence to be processed, determine the first entity corresponding to the medical entity and the node corresponding to the first entity on the knowledge graph according to the start position and end position of each medical entity in the sentence to be processed, determine the answer corresponding to the sentence to be processed by analyzing the relationship corresponding to the sentence to be processed and

the node corresponding to the first entity on the knowledge graph, and outputs the answer. Therefore, the situation that manual search is required for data from the papers corresponding to the knowledge graph in the prior art can be avoided, and thus the processing efficiency of the user's questions can be improved, which meets the user's use requirements and thus improves the user experience.

BRIEF DESCRIPTION OF THE DRAWINGS

The present disclosure will be described in more detail with reference to the accompany drawings and the embodiments, wherein in the drawings:

FIG. 1 is a schematic view showing an optional application environment of an electronic device in accordance with an embodiment of the present disclosure;

FIG. 2 is a schematic view showing a hardware structure of an electronic device in accordance with a first embodiment of the present disclosure;

FIG. 3 is a functional block diagram of the electronic device in accordance with the first embodiment of the present disclosure;

FIG. 4 is a schematic view showing a displaying effect of nodes on a knowledge graph in accordance with an embodiment of the present disclosure;

FIG. 5 is a flowchart showing a medical knowledge graph question answering processing method in according with a first embodiment of the present disclosure;

FIG. 6 is a flowchart showing a medical knowledge graph question answering processing method in according with a second embodiment of the present disclosure;

FIG. 7 is a flowchart showing a medical knowledge graph question answering processing method in according with a third embodiment of the present disclosure; and

FIG. 8 is a flowchart showing a medical knowledge graph question answering processing method in according with a fourth embodiment of the present disclosure.

PREFERRED EMBODIMENTS

For clearly understanding technical features, purpose, and effect of the present disclosure, embodiments are given in detail hereinafter with reference to the accompanying drawings. It should be understood that the specific embodiments here are used only to interpret this application and not limit to this application. All other implementation of the case obtained by one having ordinary skill in the art without creative work should be considered within the scope of this application for protection.

It is to be noted that descriptions involving “first”, “second” and the like in the disclosure are only made for description and should not be understood to indicate or imply relative importance thereof or implicitly indicate the number of indicated technical features. Therefore, a feature defined by “first” and “second” may explicitly or implicitly indicate inclusion of at least one such feature. In addition, the technical solutions of each embodiment may be combined on the premise that the combinations may be implemented by those of ordinary skill in the art, and if the combinations of the technical solutions conflict or may not be implemented, it should be understood that the combinations of the technical solutions do not exist and also not fall within the scope of protection claimed by the disclosure.

FIG. 1 is a schematic view of an optional application environment of an electronic device 20 of the present disclosure.

In an embodiment, the electronic device 20 can communicate with a terminal 11 and a database 30 in a wired or wireless manner. The electronic device 20 obtains input information from the terminal 11 through a network interface 23, retrieves corresponding knowledge graph data information from the database 30 according to the obtained input information after the input information is processed, and sends the data information to a display interface of the terminal 11 through the network interface 23, thereby realizing the transmission of data after the medical knowledge graph question answering processing. The terminal 11 includes a mobile phone, a tablet, a personal computer, and so on. The database 30 at least includes a data server.

Referring to FIG. 2, which is a schematic view showing an optional hardware structure of the electronic device 20 of the present disclosure, the electronic device 20 includes but not limited to a storage 21, a processor 22, and the network interface 23 that can communicate with each other through a system bus. FIG. 2 only shows the electronic device 20 with the components 21-23, and it should be understood that not all the components are required to be implemented; in other embodiments, more or less components may be implemented instead.

The storage 21 includes at least one type of readable storage medium. The readable storage medium includes a flash memory, a hard disk, a multimedia card, a card-type memory (for example, 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 disk, etc. In some embodiments, the storage 21 may be an internal storage unit of the electronic device 20, such as a hard disk or a memory of the electronic device 20. In other embodiments, the storage 21 may also be an external storage device of the electronic device 20, such as a plug-in hard disk equipped on the electronic device 20, a smart media card (SMC), a secure digital, a flash card, etc. Of course, the storage 21 may also include both the internal storage unit of the electronic device 20 and the external storage device. In this embodiment, the storage 21 is generally used to store an operating system and various application software installed in the electronic device 20, such as program codes of a medical knowledge graph question answering processing system 24. In addition, the storage 21 can also be used to temporarily store various types of data that have been output or will be output.

The processor 22 may be a central processing unit (CPU), a controller, a microcontroller, a microprocessor, or other data processing chips. The processor 22 is generally used to control the overall operation of the electronic device 20. In this embodiment, the processor 22 is used to run the program code stored in the storage 21 or process data, for example, to run a medical knowledge graph question answering processing system 24.

The network interface 23 may include a wireless network interface or a wired network interface, and the network interface 23 is generally used to establish a communication connection between the electronic device 20 and other electronic devices.

So far, the hardware structure and functions of the relevant equipment of the present disclosure have been introduced in detail. Hereinafter, various embodiments of the present disclosure will be presented based on

the above introduction.

First, the present disclosure provides the electronic device 20.

Referring to FIG. 3, which is a functional block diagram of the electronic device 20 according to a first embodiment of the present disclosure.

In the embodiment, the electronic device 20 includes a series of computer program instructions stored on the storage 21. When the computer program instructions are executed by the processor 22, the medical knowledge graph question answering processing operation of each embodiment of the present disclosure can be implemented. In some embodiments, the electronic device 20 may be divided into one or more modules based on specific operations implemented by the various parts of the computer program instructions. For example, in FIG. 3, the electronic device 20 may be divided into a recognition module 201, an obtaining module 202, a determining module 203, a processing module 204, and an output module 205.

The recognition module 201 is configured to receive a sentence to be processed sent to the electronic device from external operation.

In an embodiment, the recognition module 201 receives the sentence to be processed and recognizes a medical entity in the sentence to be processed.

In order to further improve the efficiency of processing the sentence to be processed, avoid a situation that the sentence sent by users does not meet a trigger condition for the medical knowledge graph question answering processing operation, and thus to improve the efficiency of processing the sentence to be processed, in an embodiment, the obtaining of the sentence to be processed includes: receiving a sentence sent by a user

and determining whether the sentence sent by the user is a question sentence; if the sentence sent by the user is a question sentence, determining that the sentence of sent by the user is the sentence to be processed.

It is understood that when the sentence is not a question sentence, it means that the sentence does not need to be answered. Therefore, by directly judging whether the sentence is a question sentence or not, whether sentence processing is really necessary or not can be determined, which improves the pertinence of sentence processing and avoids the processing of invalid sentences, thereby improving the processing efficiency of the sentence to be processed as a whole.

Those skilled in the art can understand that one or more medical entities can be recognized. Because the medical entity is mainly used in the medical field, the recognized medical entity may be a name of a disease, a name of a medicine, and so on.

It is understood that the sentence to be processed is a question sent by the user that needs to be solved. For example, the sentence to be processed is: "How should atorvastatin be taken to prevent coronary heart disease"? The medical entities obtained after recognition are: atorvastatin (drug), coronary heart disease (disease).

In an embodiment, the recognition module 201 is configured to obtain the sentence to be processed, and recognize the medical entity in the sentence to be processed using a NER model, wherein the medical entity at least includes: diseases and/or medicines.

It should be noted that training data of the NER model (named entity recognition model, NER model for short) is formed by manually labeling the collected question sentence data set. The NER model uses the popular bi-LSTM-CRF model; the input of the model is an embedding layer based

on Chinese characters and an embedding layer based on radicals, and the added features of the model include a part-of-speech labeling in the question sentence and a word category labeling in the question sentence by loading a medical dictionary according to a stammering tool. A mature NER model can be obtained by training the model with training samples in the medical field and then the mature NER model can be used to recognize the medical entity. For those skilled in the art, this part is a conventional method and will not be repeated in embodiments of the present disclosure hereinafter.

Therefore, the entity recognition method formed by the recognition network structure and the feature recognition of the embodiment of the present disclosure can assist in effectively recognizing the entity in the sentence to be processed.

The obtaining module 202 is used to obtain a start position and an end position of each medical entity in the sentence to be processed.

It should be noted that after the medical entity in the sentence to be processed is obtained, a specific position of the medical entity in the sentence to be processed can be obtained for each medical entity, and the specific position includes the start position and the end position.

Exemplarily, the start position of the medical entity “atorvastatin” recognized from the sentence to be processed “How should atorvastatin be taken to prevent and treat coronary heart disease?” is the 1st word, and the end position thereof is the 5th word. The start position of the medical entity “coronary heart disease” is the 8th word, and the end position thereof is the 10th word.

The determining module 203 is configured to compare the medical entity with entities in a preset knowledge base according to the medical

entity and the corresponding start position and end position to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph.

Because a drug may have multiple names, or a same type of drug for treating a certain disease may be produced by multiple manufacturers, the situation that the medical entity recognized by the recognition module 201 is directly used may cause too much limitation, and thus a comprehensive result cannot be obtained. For example, for the sentence to be processed “the dosage of cold medicine?”, the medical entities are cold medicines (drugs) and colds (disease); since cold medicines produced by different manufacturers are taken differently, for example, the unit of a granular cold medicine is a bag and the unit of a capsule cold medicine is a capsule, therefore, the comparison of the medical entity in the knowledge base is required to determine the entity corresponding to the medical entity “cold medicine” in the knowledge base, that is, the first entity. It is assumed that the obtained first entities include: adult cold medicine and children cold medicine.

It is understood that the adult cold medicine in the first entity has the corresponding node on the knowledge graph, and the children cold medicine in the first entity also has the corresponding node on the knowledge graph. After the node is found on the knowledge graph, the content contained in the node can be obtained through the node. For example, the nodes corresponding to the adult cold medicine may correspond to the traditional Chinese cold medicine and the western cold medicine, including the name and dosage of the traditional Chinese cold medicine, and the name and dosage of the western cold medicine. In this way, it is more detailed and more comprehensive than obtaining the node division on the knowledge

graph directly through cold medicine.

[0001] In an embodiment, a form of each medical entity in the sentence to be processed is determined according to the medical entity and the corresponding start position and end position; the form of each medical entity in the sentence to be processed is compared with the entities in the knowledge base to determine a second entity corresponding to the medical entity; according to a similarity algorithm, a similarity between the medical entity and the second entity is determined; and according to the determined similarity, the first entity matching the medical entity is determined from the second entity.

It should be noted that by analyzing the question sentence, the expression form of each medical entity in the sentence to be processed can be obtained. There are many question sentences stored in the knowledge base each of which also contains entities, and the expression form of each entity also can be obtained by analyzing the entity. The expression form of the entity in the sentence to be processed is compared with the entity in the knowledge base to determine whether the medical entity in the sentence to be processed has other names, thereby obtaining more comprehensive medical entities. The more medical entities obtained, the greater the possibility of errors. Therefore, in order to improve the accuracy of the medical entity, the similarity between the medical entity and the second entity is determined to obtain more accurate first entities corresponding to the medical entity.

In an embodiment, those skilled in the art can understand that the similarity can be calculated using a Chinese character editing distance, a Pinyin editing distance, a Word2vec feature, and a synonym thesaurus feature. Therefore, in an embodiment, a similarity value between the

medical entity and the second entity is calculated according to an editing distance ratio of character string; and the similarity between the medical entity and the second entity is determined according to a magnitude of the similarity value; or, the similarity value between each medical entity and the second entity is calculated according to the feature vector, and the similarity between the medical entity and the second entity is determined according to a magnitude of the similarity value.

In an embodiment, the calculated similarity value can be compared with a preset similarity value. When the similarity value is less than the preset similarity value, the corresponding second entity is deleted and not used; otherwise, when the second entity is greater than the preset similarity value, the second entity is determined to be the first entity.

By calculating the similarity between the first entity and the medical entity, irrelevant entities can be eliminated from the first entity, thereby obtaining more accurate second entity. It is understood that the number of entities contained in the first entity is often larger than that of entities contained in the second entity. In some embodiments, the calculation result may also indicate that the first entity is highly relative with the medical entity, in this case, the first entity and the second entity can be the same entity set.

Exemplarily, the first medical entity corresponding to the medical entity “atorvastatin” includes: “atorvastatin calcium tablets”, “drug”, “atorvastatin calcium dispersible tablets”, “drug”, “atorvastatin calcium capsules”, “drug”, “amlodipine atorvastatin calcium tablets” “drug”, “atorvastatin calcium”, “ATC”, “rosuvastatin calcium tablets”, “drug”. Therefore, after the similarity value is calculated, the similarity between “atorvastatin calcium” and “atorvastatin” and the similarity between “ATC”

and “atorvastatin” are lower than the preset similarity value, thus, “Atorvastatin calcium” and “ATC” are eliminated from the first entity, and the other entities in the first entity are used as the second entity.

Therefore, the entities can be effectively recognized and noises can be filtered by adopting boundary adjustment and multi-dimensional similarity calculation in the entity connection part.

The processing module 204 is configured to perform relationship analysis on the sentence to be processed, and obtain a relationship corresponding to the sentence to be processed based on a relationship matching model.

Exemplarily, the relationship of the sentence to be processed “how should atorvastatin be taken to prevent coronary heart disease?” is: “How should <drugs> prevent and treat <disease>?”. Both <drug> and <disease> represent medical entities.

In one embodiment, the training process of the relationship matching model includes: collecting samples, and manually labeling the samples as positive sample question sentences and negative sample question sentences, wherein the positive sample question sentence is an entity having a corresponding relationship in the knowledge graph, and the negative sample question sentence is an entity that does not have a corresponding relationship in the knowledge graph; training a LSTM network with the positive sample question sentences and the negative sample question sentences; according to a training output value, determining a maturity of the relationship matching model; and using the LSTM having a maturity greater than the preset value as the relationship matching model.

It is understood that, for each question sentence, if the question sentence has the corresponding relationship in the knowledge graph, the

question sentence is a positive sample question sentence; otherwise the question sentence is a negative sample question sentence. The ratio of positive and negative sample question sentences is determined according to the number of relationships in the knowledge graph, for example, the ratio can be 1:50. The relationship matching model adopts the improved Enhanced Log-Structured Merge Tree, referred to as ESIM network. This network structure is originally mainly used for question sentence pair matching tasks. The specific process is: the question sentence input by users enters an embedding layer, and then passes through a two-way LSTM network for encoding and a relationship embedding layer which is composed of two parts: (a) random initialization vector; (b) a vector output after a relationship name passes through the embedding layer and the LSTM model. After interacting with the attention mechanism, the output passes through the LSTM network, and the final output passes through a feedforward network to score the output matching degree.

The output module 205 is configured to determine an answer corresponding to the sentence to be processed based on the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, and output the answer.

It is understood that each entity has a corresponding node content on the knowledge graph. In an embodiment, the output module 205 is configured to search the content corresponding to the relationship of the sentence to be processed on the node corresponding to the first entity on the knowledge base, determine the found content as the answer corresponding to the sentence to be processed, and output the answer.

As shown in FIG. 4, atorvastatin calcium is the main component of Lipitor. The specification of Lipitor includes 10 mg, 20 mg and 40 mg,

which are suitable for coronary heart disease. At the same time, at the node of the entity of coronary heart disease, it can be seen that both Plavix and Aspirin are adapted to the symptoms of coronary heart disease. Therefore, the corresponding answer can be: “The commonly-used starting dose of Lipitor is 10 mg once a day”. For Lipitor with a specification of 20 mg, the corresponding answer can also be output. Therefore, the output answer is more comprehensive, and can provide a relationship matching result with high accuracy and high recall rate.

Therefore, in summary, the medical knowledge graph question answering processing device provided in the present disclosure recognizes the medical entity in the sentence to be processed, determines the first entity corresponding to the medical entity and the node corresponding to the first entity on the knowledge graph according to the start position and end position of each medical entity in the sentence to be processed, determines the answer corresponding to the sentence to be processed by analyzing the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, and outputs the answer. Thus, the situation that manual search is required for data from the papers corresponding to the knowledge graph in the prior art can be avoided, and therefore the processing efficiency of the user's questions can be improved, which meets the user's use requirements and thus improves the user experience.

In addition, the present disclosure further provides a medical knowledge graph question answering processing method.

Refer to FIG. 5, which is a schematic flowchart of a medical knowledge graph question answering processing method according to a first embodiment of the present disclosure. The method is applied to the

electronic device 20. In this embodiment, according to different requirements, the execution order of the steps in the flowchart shown in FIG. 5 can be changed, and some steps can be omitted. The method includes steps as follows.

Step S501, obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

step S502, obtaining a start position and an end position of each medical entity in the sentence to be processed;

step S503, comparing the medical entity with entities in a preset knowledge base according to the medical entity and the corresponding start position and end position to determine a first entity corresponding to the medical entity and a node on the knowledge graph corresponding to the first entity;

step S504, performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model;

step S505, determining an answer corresponding to the sentence to be processed according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, and outputting the answer.

As shown in FIG. 6, the step S501, which in an embodiment is the step S601, includes: obtaining the sentence to be processed, and recognizing the medical entity in the sentence to be processed using a NER model, wherein the medical entity at least includes: disease and/or drug.

In an embodiment, the step of obtaining a sentence to be processed includes: receiving a sentence sent by a user and determining whether the sentence sent by the user is a question sentence; if so, determining that the

sentence sent by the user is the sentence to be processed.

As shown in FIG. 7, the step S503 includes:

S701, determining a form of each medical entity in the sentence to be processed according to the medical entity and the corresponding start position and end position;

S702, comparing the form of each medical entity in the sentence to be processed with the entities in the knowledge base to determine a second entity corresponding to the medical entity;

S703, determining a similarity between the medical entity and the second entity according to a similarity algorithm;

S704, according to the determined similarity, determining a first entity matching the medical entity from the second entities.

In an embodiment, step S703 includes: calculating a similarity value between the medical entity and the second entity according to an editing distance ratio of character string; and determining the similarity between the medical entity and the second entity according to a magnitude of the similarity value; or, calculating a similarity value between each medical entity and the second entity according to a feature vector, and determining the similarity between the medical entity and the second entity according to a magnitude of the similarity value.

Step S505 includes: searching a content corresponding to the relationship of the sentence to be processed in the node corresponding to the first entity on the knowledge graph, determining the found content as the answer corresponding to the sentence to be processed, and outputting the answer.

As shown in FIG. 8, a training process of the relationship matching model includes following steps:

S801, collecting samples and manually labeling the samples as positive sample question sentences and negative sample question sentences, wherein the positive sample question is an entity having a corresponding relationship with the knowledge graph, and the negative sample question is an entity that does not have a corresponding relationship with the knowledge graph;

S802, training a LSTM network using the positive sample question sentences and the negative sample question sentences;

S803, determining a maturity of a relationship matching model according to a training output value; and

S804, using the LSTM network with a maturity greater than a preset value as the relationship matching model.

The present disclosure also provides a computer apparatus, such as a smart phone, a tablet computer, a notebook computer, a desktop computer, a rack server, a blade server, a tower server or a cabinet server (including independent servers, or more Server cluster composed of multiple servers). The computer apparatus of the embodiment at least includes but not limited to a storage, a processor, and a computer program stored in the storage and running on the processor that can be communicatively connected to each other through a system bus, wherein when the processor executes the computer program, a medical knowledge graph question answering processing method is implemented, and the method includes the following steps:

obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

obtaining a start position and an end position of each medical entity in the sentence to be processed;

according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity, and a node corresponding to the first entity on the knowledge graph;

performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model;

according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

The present disclosure also provides a computer-readable storage medium. The computer-readable storage medium may be a nonvolatile storage medium or a volatile storage medium, such as flash memory, hard disk, multimedia card, card-type memory (for example, SD or DX). Memory, etc.), random access memory (RAM), static random access memory (SRAM), read only memory (ROM), electrically erasable programmable read only memory (EEPROM), programmable read only memory (PROM), magnetic memory, magnetic disk, optical disk, server, App application mall, etc. Computer programs are stored on the storage medium, and the corresponding functions are performed when the programs are executed by a processor. The computer-readable storage medium of this embodiment is used to store the electronic device 20 which, when being executed by a processor, performs a medical knowledge graph question answering processing method of the present disclosure. The method includes the following steps:

obtaining a sentence to be processed, and recognizing a medical entity

in the sentence to be processed;

obtaining a start position and an end position of each medical entity in the sentence to be processed;

according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity, and a corresponding node corresponding to the first entity on the knowledge graph;

performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model;

according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

The serial numbers of the foregoing embodiments of the present application are only for description, and do not represent the advantages and disadvantages of the embodiments.

Through the description of the above embodiments, those skilled in the art can clearly understand that the method of the above embodiments can be implemented by means of software plus the necessary general hardware platform. Of course, it can also be implemented by hardware, but in many cases the former is the better embodiment. Based on this understanding, the technical solution of this application essentially or the part that contributes to the existing technology can be embodied in the form of a software product, and the computer software product is stored in a storage medium (such as ROM/RAM, magnetic disk, The optical disc)

includes several instructions to enable a terminal (which may be a mobile phone, a computer, a server, an air conditioner, or a network device, etc.) to execute the method described in each embodiment of the present application.

The foregoing embodiments are merely intended for describing the technical solutions of the present invention rather than limiting the present invention. Although the present invention is described in detail with reference to the foregoing embodiments, persons of ordinary skill in the art should understand that they may still make modifications to the technical solutions described in the foregoing embodiments, or make equivalent replacements to some technical features thereof, as long as such modifications or replacements do not cause the essence of corresponding technical solutions to depart from the spirit and scope of the technical solutions of the embodiments of the present invention.

WHAT IS CLAIMED IS:

1. A medical knowledge graph question answering processing method, comprising:

obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

obtaining a start position and an end position of each medical entity in the sentence to be processed;

according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model; and

according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

2. The medical knowledge graph question answering processing method according to claim 1, wherein the step of obtaining a sentence to be processed and recognizing a medical entity in the sentence to be processed comprises: obtaining the sentence to be processed, and recognizing the medical entity in the sentence to be processed using a NER model, wherein the medical entity at least comprises: disease and/or medicine.

3. The medical knowledge graph question answering processing method according to claim 1 or 2, wherein the step of obtaining a sentence

to be processed comprises: receiving a sentence sent by an user, and determining whether the sentence sent by the user is a question sentence; if the sentence sent by the user is a question sentence, determining that the sentence sent by the user is the sentence to be processed.

4. The medical knowledge graph question answering processing method according to claim 3, wherein the step of according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph comprises:

- determining a form of each medical entity in the sentence to be processed according to the medical entity and the corresponding start position and end position;

- comparing the form of each medical entity in the sentence to be processed with the entities in the knowledge base to determine a second entity corresponding to the medical entity;

- determining a similarity between the medical entity and the second entity according to a similarity algorithm; and

- according to the determined similarity, determining the first entity matching the medical entity from the second entity.

5. The medical knowledge graph question answering processing method according to claim 4, wherein the step of determining a similarity between the medical entity and the second entity according to a similarity algorithm comprises:

- calculating a similarity value between the medical entity and the second entity according to an editing distance ratio of character string;

- determining the similarity between the medical entity and the second

entity according to a magnitude of the similarity value; or,

calculating a similarity value between each medical entity and the second entity according to a feature vector;

determining the similarity between the medical entity and the second entity according to a magnitude of the similarity value.

6. The medical knowledge graph question answering processing method according to claim 1, wherein the step of according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer comprises: searching a content corresponding to the relationship of the sentence to be processed in the node corresponding to the first entity on the knowledge graph; determining the found content as the answer corresponding to the sentence to be processed, and outputting the answer.

7. The medical knowledge graph question answering processing method according to claim 1, wherein a training step of the relationship matching model comprises:

collecting samples, and manually labeling the samples as positive sample question sentences and negative sample question sentences, wherein the positive sample question sentence is an entity that has a corresponding relationship with the knowledge graph, and the negative sample question is an entity that does not have a corresponding relationship with the knowledge graph;

training a LSTM network using the positive sample question sentences and the negative sample question sentences;

determining a maturity of the relationship matching model according to a training output value; and

using the LSTM network with a maturity greater than a preset value as the relational matching model.

8. An electronic device, comprising:

a recognition module configured to obtain a sentence to be processed and to recognize a medical entity in the sentence to be processed;

an obtaining module configured to obtain a start position and an end position of each medical entity in the sentence to be processed;

a determining module configured to, according to the medical entity and the corresponding start position and end position, compare the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

a processing module configured to perform relationship analysis on the sentence to be processed, and obtain a relationship corresponding to the sentence to be processed based on a relationship matching model; and

an output module configured to determine an answer corresponding to the sentence to be processed according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, and output the answer.

9. An apparatus comprising a storage, a processor, and a computer program stored in the storage and capable of running on the processor, wherein when the processor implements the computer program, a medical knowledge graph question answering processing method is performed, and the method comprises the following steps:

obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

obtaining a start position and an end position of each medical entity in

the sentence to be processed;

according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model; and

according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

10. The apparatus according to claim 9, wherein the step of obtaining a sentence to be processed and recognizing a medical entity in the sentence to be processed comprises: obtaining the sentence to be processed, and recognizing the medical entity in the sentence to be processed using a NER model, wherein the medical entity at least comprises: disease and/or medicine.

11. The apparatus according to claim 9 or 10, wherein the step of obtaining a sentence to be processed comprises: receiving a sentence sent by an user, and determining whether the sentence sent by the user is a question sentence; if the sentence sent by the user is a question sentence, determining that the sentence sent by the user is the sentence to be processed.

12. The apparatus according to claim 11, wherein the step of according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node

corresponding to the first entity on the knowledge graph comprises:

- determining a form of each medical entity in the sentence to be processed according to the medical entity and the corresponding start position and end position;

- comparing the form of each medical entity in the sentence to be processed with the entities in the knowledge base to determine a second entity corresponding to the medical entity;

- determining a similarity between the medical entity and the second entity according to a similarity algorithm; and

- according to the determined similarity, determining the first entity matching the medical entity from the second entity.

13. The apparatus according to claim 12, wherein the step of determining a similarity between the medical entity and the second entity according to a similarity algorithm comprises:

- calculating a similarity value between the medical entity and the second entity according to an editing distance ratio of character string;

- determining the similarity between the medical entity and the second entity according to a magnitude of the similarity value; or,

- according to a feature vector, calculating a similarity value between each medical entity and the second entity; and

- determining the similarity between the medical entity and the second entity according to a magnitude of the similarity value.

14. The apparatus according to claim 9, wherein the step of according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer comprises:

searching a content corresponding to the relationship of the sentence to be processed in the node corresponding to the first entity on the knowledge graph; determining the found content as the answer corresponding to the sentence to be processed, and outputting the answer.

15. A computer-readable storage medium on which a computer program is stored, wherein when the computer program is executed by a processor, a medical knowledge graph question answering processing method is implemented, comprising the following steps:

obtaining a sentence to be processed, and recognizing a medical entity in the sentence to be processed;

obtaining a start position and an end position of each medical entity in the sentence to be processed;

according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph;

performing relationship analysis on the sentence to be processed, and obtaining a relationship corresponding to the sentence to be processed based on a relationship matching model; and

according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer.

16. The computer-readable storage medium according to claim 15, wherein the step of obtaining a sentence to be processed and recognizing a medical entity in the sentence to be processed comprises: obtaining the sentence to be processed, and recognizing the medical entity in the sentence

to be processed using a NER model, wherein the medical entity at least comprises: disease and/or medicine.

17. The computer-readable storage medium according to claim 15 or 16, wherein the step of obtaining sentences to be processed comprises: receiving a sentence sent by an user, and determining whether the sentence sent by the user is a question sentence; if the sentence sent by the user is a question sentence, determining that the sentence sent by the user is the sentence to be processed.

18. The computer-readable storage medium according to claim 17, wherein the step of according to the medical entity and the corresponding start position and end position, comparing the medical entity with entities in a preset knowledge base to determine a first entity corresponding to the medical entity and a node corresponding to the first entity on the knowledge graph comprises:

- determining a form of each medical entity in the sentence to be processed according to the medical entity and the corresponding start position and end position;

- comparing the form of each medical entity in the sentence to be processed with the entities in the knowledge base to determine a second entity corresponding to the medical entity;

- determining a similarity between the medical entity and the second entity according to a similarity algorithm; and

- according to the determined similarity, determining the first entity matching the medical entity from the second entity.

19. The computer-readable storage medium according to claim 18, the step of determining a similarity between the medical entity and the second entity according to a similarity algorithm comprises:

calculating a similarity value between the medical entity and the second entity according to an editing distance ratio of character string;

determining the similarity between the medical entity and the second entity according to a magnitude of the similarity value; or,

according to a feature vector, calculating a similarity value between each medical entity and the second entity; and

determining the similarity between the medical entity and the second entity according to a magnitude of the similarity value.

20. The computer-readable storage medium according to claim 15, wherein the step of according to the relationship corresponding to the sentence to be processed and the node corresponding to the first entity on the knowledge graph, determining an answer corresponding to the sentence to be processed and outputting the answer comprises:

searching a content corresponding to the relationship of the sentence to be processed in the node corresponding to the first entity on the knowledge graph; determining the found content as the answer corresponding to the sentence to be processed, and outputting the answer.