INFORMATION RETRIEVAL METHOD, DEVICE, AND STORAGE MEDIUM

TECHNICAL FIELD

[0001] This specification relates to the field of Internet technologies, and in particular, to an information retrieval method, a device, and a storage medium.

BACKGROUND

[0002] With rapid development of information technology, an explosive growth of a data amount makes an information retrieval technology an indispensable part of a modern society. Regardless of academic research or personal life, required information is obtained based on the information retrieval technology. Some retrieval platforms provide users with retrieval manners for information retrieval in a document library. The document library can include a plurality of documents. The user can input a retrieval requirement to the retrieval platform, for example, a retrieval keyword or a question. The retrieval platform retrieves related content from the document library based on the retrieval requirement, and generates a retrieval result. As a data amount in the document library continuously increases, how to improve efficiency and accuracy of information retrieval is an urgent problem to be resolved at present.

[0003] Content of the background part is merely information learned of by the inventor, and neither means that the information has entered a public domain before an application date of the present disclosure, nor means that the information can become a conventional technology of the present disclosure.

SUMMARY

[0004] This specification provides an information retrieval method, a device, and a storage medium, to improve retrieval efficiency and accuracy.

[0005] According to a first aspect, this specification provides an information retrieval method, including: receiving a retrieval request for requesting to perform information retrieval in a

document library, where the retrieval request includes retrieval description information, the document library includes at least one document, each document includes a level title of at least one level, text content corresponding to at least some level titles includes at least one document block, and at least some document blocks include at least one keyword; determining, in the document library based on a knowledge map corresponding to the document library, at least one target document block associated with the retrieval description information, where the knowledge map represents at least a belonging relationship between level titles of different levels, a belonging relationship between the document block and the level title, and a belonging relationship between the keyword and the document block; and generating a retrieval result corresponding to the retrieval request based on the at least one target document block.

[0006] According to a second aspect, this specification further provides an electronic device, including: at least one storage medium, storing at least one instruction set, to perform information retrieval; and at least one processor, communicatively connected to the at least one storage medium. When running, the at least one processor reads the at least one instruction set, and performs the following operations based on an indication of the at least one instruction set: receiving a retrieval request for requesting to perform information retrieval in a document library, where the retrieval request includes retrieval description information, the document library includes at least one document, each document includes a level title of at least one level, text content corresponding to at least some level titles includes at least one document block, and at least some document blocks include at least one keyword; determining, in the document library based on a knowledge map corresponding to the document library, at least one target document block associated with the retrieval description information, where the knowledge map represents at least a belonging relationship between level titles of different levels, a belonging relationship between the document block and the level title, and a belonging relationship between the keyword and the document block; and generating a retrieval result corresponding to the retrieval request based on the at least one target document block.

[0007] According to a third aspect, this specification provides a computer-readable non-transitory storage medium. The computer-readable non-transitory storage medium stores at least one instruction set, and when the at least one instruction set is executed by at least one processor, the information retrieval method according to the first aspect is implemented.

[0008] Other functions of the information retrieval method, the device, and the storage

medium provided in this specification are partially listed in the following descriptions. Creative aspects of the information retrieval method, the device, and the storage medium provided in this specification can be fully explained by practice or by using the methods, apparatuses, and combinations described in the following detailed examples.

BRIEF DESCRIPTION OF DRAWINGS

[0009] To describe the technical solutions in the embodiments of this specification more clearly, the following briefly describes the accompanying drawings needed for describing the embodiments. Clearly, the accompanying drawings in the following descriptions show merely some embodiments of this specification, and a person of ordinary skill in the art can still derive other drawings from these accompanying drawings without creative efforts.

[0010] FIG. 1 is a schematic diagram of an application scenario of information retrieval according to an embodiment of this specification;

[0011] FIG. 2 is a diagram of a hardware structure of an electronic device according to an embodiment of this specification;

[0012] FIG. 3 is a flowchart of an information retrieval method according to an embodiment of this specification;

[0013] FIG. 4 is a schematic diagram of a document according to an embodiment of this specification;

[0014] FIG. 5 is a schematic diagram of another document according to an embodiment of this specification;

[0015] FIG. 6 is a schematic diagram of a knowledge map according to an embodiment of this specification;

[0016] FIG. 7 is another schematic diagram of a knowledge map according to an embodiment of this specification;

[0017] FIG. 8 is a schematic flowchart of generating a knowledge map according to an embodiment of this specification; and

[0018] FIG. 9 is a schematic diagram of applying an information retrieval method to a large model according to an embodiment of this specification.

DESCRIPTION OF EMBODIMENTS

[0019] Specific application scenarios and requirements of this specification are provided in the following descriptions, so that a person skilled in the art can manufacture and use content of this specification. For the person skilled in the art, various local modifications to the disclosed embodiments are clear, and the general principles defined here can be applied to other embodiments and applications without departing from the spirit and scope of this specification. Therefore, this specification is not limited to the shown embodiments, but has a widest scope consistent with that of the claims.

[0020] The terms used here are only intended to describe a particular example embodiment, but do not impose a limitation. For example, unless otherwise specified in the context, the singular forms "one", "a", and "the" used here can also include a plural form. When used in this specification, the terms "include", "comprise", and/or "have" indicate/indicates existence of an associated integer, step, operation, element, and/or component, but does not exclude existence of one or more other features, integers, steps, operations, elements, components, and/or groups, or another feature, integer, step, operation, element, component, and/or group can be added to the system/method.

[0021] In consideration of the following descriptions, these and other features of this specification, operations and functions of related components of the structure, and economy of combination and manufacturing of components can be significantly improved. With reference to the accompanying drawings, all of these form a part of this specification. However, it should be clearly understood that the accompanying drawings are merely used for the purpose of illustration and description, and are not intended to limit the scope of this specification. It should be further understood that the accompanying drawings are not drawn to scale.

[0022] The flowchart used in this specification illustrates operations implemented by a system in some embodiments of this specification. It should be clearly understood that operations in the flowchart can be implemented out of sequence. On the contrary, operations can be implemented reversely or simultaneously. In addition, one or more other operations can be added to the flowchart. One or more operations can be removed from the flowchart.

[0023] In this specification, a retrieval platform can be a system that provides an information retrieval service. The retrieval platform can also be referred to as a retrieval system. In some

embodiments, the retrieval platform can be an independent platform. For example, the retrieval platform directly provides a retrieval service to a user, and the user inputs a retrieval requirement into the retrieval platform. After information retrieval, the retrieval platform returns a retrieval result to the user. In some embodiments, the retrieval platform can alternatively be not directly user-oriented, but serves as a module or unit of another service platform, and in a process in which the another service platform executes a service, can invoke an information retrieval service provided by the retrieval platform.

[0024] In some embodiments, the retrieval platform has a function of information retrieval in a document library. The document library includes at least one document, and each document includes at least one document paragraph. In a process in which the user performs information retrieval on the retrieval platform, the retrieval platform can receive retrieval description information (for example, a retrieval keyword or a question) entered by the user. Therefore, the retrieval platform can obtain a target document paragraph corresponding to the retrieval description information from the document library based on the retrieval description information, and generate the retrieval result based on the target document paragraph.

[0025] For example, the retrieval platform quantizes the retrieval description information, to obtain a vector representation of the retrieval description information, and determines, based on a distance between the vector representation of the retrieval description information and a vector representation of each document paragraph in the document library, a target document paragraph related to a retrieval requirement. In a process of generating the retrieval result, the retrieval platform can invoke a large model to understand the target document paragraph, and provide the retrieval result corresponding to the retrieval description information.

[0026] In the above-mentioned embodiment, when determining the target document paragraph, the retrieval platform can calculate a distance between the vector representation of the retrieval description information and a vector of each word segment in a specific document paragraph, to determine whether the document paragraph serves as the target document paragraph. Because each document paragraph includes a relatively large quantity of word segments, a relatively long time is consumed in a process in which the retrieval platform calculates a distance through traversal, and consequently, retrieval efficiency is not high. In addition, if similar content exists in different documents, the retrieval platform easily determines that all of a plurality of document paragraphs that are superficially similar and that have different substances are the target document paragraph,

and further confuses similar content in a process of generating the retrieval result, resulting in an inaccurate retrieval result.

[0027] In view of this, this specification provides an information retrieval method, a device, and a storage medium, to improve efficiency and accuracy of information retrieval.

[0028] The following describes an application scenario of information retrieval provided in this specification with reference to FIG. 1.

[0029] FIG. 1 is a schematic diagram of an application scenario of information retrieval according to an embodiment of this specification. As shown in FIG. 1, a scenario 001 can include a user 100, a terminal device 200, a server 300, a network 400.

[0030] The user 100 can execute a related event by using the terminal device 200. For example, the terminal device 200 is installed with a client of one or more service platforms, and the client can provide the user 100 with a capability of executing the related event. The service platform includes but is not limited to a web browser platform, a retrieval platform, a chat platform, a shopping platform, a service platform, a video platform, a financial management platform, or a payment platform. For example, the terminal device 200 can present an interaction interface to the user, so that the user 100 performs information retrieval in the interaction interface. The user 100 is, for example, a retrieval person.

[0031] In some embodiments, the terminal device 200 can perform the information retrieval method described in this specification. The terminal device 200 can store data and instructions used to implement the information retrieval method, and can execute or be configured to execute the data and instructions. In some embodiments, the terminal device 200 can include a hardware device with a data information processing function and a necessary program needed for driving the hardware device to work.

[0032] For example, the terminal device 200 receives a retrieval request for requesting to perform information retrieval in a document library, where the retrieval request includes retrieval description information; determines, in the document library based on a knowledge map corresponding to the document library, at least one target document block related to the retrieval description information; and further generates a retrieval result corresponding to the retrieval request based on the at least one target document block. The terminal device 200 can display the retrieval result to the user 100. An implementation of steps in this example and related terms involved are described in detail in the following method P100.

[0033] The terminal device 200 can include a mobile device, a tablet, a laptop computer, a device built in a motor vehicle or similar content, or any combination thereof. In some embodiments, the mobile device can include a smart home device, a smart mobile device, a virtual reality device, an augmented reality device or a similar device, or any combination thereof. In some embodiments, the smart home apparatus can include a smart television, a desktop computer, or any combination. In some embodiments, the smart mobile device can include a smartphone, a personal digital assistant, a game device, a navigation device, or any combination thereof. In some embodiments, the virtual reality device or the augmented reality device may include a virtual reality helmet, virtual reality glasses, a virtual reality patch, an augmented reality helmet, augmented reality glasses, an augmented reality patch or similar content, or any combination thereof. For example, the virtual reality device or the augmented reality device may include AR glasses, a head-mounted display, and VR. In some embodiments, the apparatus built in the motor vehicle can include a vehicle-mounted computer, a vehicle-mounted television, etc. In some embodiments, the terminal device 200 can be a device with a location technology, and is configured to locate a location of the terminal device 200.

[0034] As shown in FIG. 1, the terminal device 200 can be communicatively connected to the server 300. The server 300 can be a server that provides various services, for example, a back-end server that provides an information retrieval service for the terminal device 200.

[0035] In some embodiments, the server 300 can perform the information retrieval method described in this specification. The server 300 can store data and instructions for implementing then information retrieval method, and can execute or be configured to execute the data and instructions. In some embodiments, the server 300 can include a hardware device with a data information processing function and a necessary program needed for driving the hardware device to work.

[0036] For example, after obtaining retrieval description information entered by the user, the terminal device 200 can generate a retrieval request that includes the retrieval description information, and send the retrieval request to the server 300. After receiving the retrieval request, the server 300 determines, in the document library based on the knowledge map corresponding to the document library, at least one target document block related to the retrieval description information, and further generates a retrieval result corresponding to the retrieval request based on the at least one target document block. The server 300 can send the retrieval result to the terminal

device 200. The terminal device 200 can display the retrieval result to the user 100. An implementation of steps in this example and related terms involved are described in detail in the following method P100.

[0037] The network 400 is a medium configured to provide a communication connection between the terminal device 200 and the server 300. The network 400 can facilitate exchange of information or data. In some embodiments, the network 400 can be any type of wired or wireless network, or a combination thereof. For example, the network 400 can include a cable network, a wired network, an optical fiber network, a telecommunications network, an intranet, the Internet, a local area network (LAN), a wide area network (WAN), a wireless local area network (WLAN), a metropolitan area network (MAN), a public switched telephone network (PSTN), a Bluetooth network, a ZigBee network, a near field communication (NFC) network, or a similar network. In some embodiments, the network 400 can include one or more network access points. For example, the network 400 can include a wired or wireless network access point, for example, a base station or an Internet switching point. Through the access point, one or more components in the terminal device 200 and the server 300 can be connected to the network 400 to exchange data or information.

[0038] It should be understood that quantities of terminal devices 200, servers 300, and networks 400 in FIG. 1 are merely examples. Based on an implementation requirement, there can be any quantity of terminal devices 200, servers 300, and networks 400.

[0039] It should be noted that the information retrieval method can be completely performed on the terminal device 200, or can be completely performed on the server 300, or can be partially performed on the terminal device 200 and partially performed on the server 300. This is not limited in this embodiment of this specification.

[0040] It should be noted that FIG. 1 shows a case in which the retrieval platform directly provides the information retrieval service to the user. To be specific, the retrieval platform includes a client and a serving end, the client is deployed on the terminal device 200, and the serving end is deployed on the server 300. In some embodiments, the retrieval platform can alternatively provide the information retrieval service to another service system instead of directly providing the information retrieval service to the user. In this case, in a running process of the another service system, when a function of performing information retrieval in the document library needs to be implemented, the another service system can send the retrieval request to the retrieval platform. The retrieval platform performs, in response to the retrieval request, the information retrieval

method provided in this specification, to obtain the retrieval result, and feeds back the retrieval result to the another service system, so that the another service system performs subsequent processing based on the retrieval result.

[0041] FIG. 2 is a diagram of a hardware structure of an electronic device 600 according to an embodiment of this specification. The electronic device 600 can be a device on which a retrieval platform is deployed. In some embodiments, the electronic device 600 can serve as the terminal device 200 or the server 300 in FIG. 1.

[0042] As shown in FIG. 2, the electronic device 600 can include at least one storage medium 630 and at least one processor 620. In some embodiments, the electronic device 600 can further include a communication port 650 and an internal communication bus 610. In addition, the electronic device 600 can further include an I/O component 660.

[0043] The internal communication bus 610 can be connected to different system components. For example, the internal communication bus 610 can be connected to the storage medium 630, the processor 620, the communication port 650, and the I/O component 660.

[0044] The I/O component 660 supports an input/output between the electronic device 600 and another component.

[0045] The communication port 650 is used for data communication between the electronic device 600 and the outside. For example, the communication port 650 can be used for data communication between the electronic device 600 and a network 400. The communication port 650 can be a wired communication port, or can be a wireless communication port.

[0046] The storage medium 630 can include a data storage apparatus. The data storage apparatus can be a non-transitory storage medium, or can be a transitory storage medium. For example, the data storage apparatus can include one or more of a disk 632, a read-only storage medium (ROM) 634, or a random access storage medium (RAM) 636. The storage medium 630 further includes at least one instruction set stored in the data storage apparatus. The instruction set can include computer program code, and the computer program code can include a program, a routine, an object, a component, a data structure, a process, a module, etc. for performing the information retrieval method provided in this specification.

[0047] The processor 620 can be communicatively connected to the storage medium 630. The processor 620 is configured to execute the at least one instruction set. When the electronic device 600 runs, the processor 620 reads the at least one instruction set, and performs, based on an

indication of the at least one instruction set, the information retrieval method provided in this specification.

[0048] The processor 620 can be in a form of one or more processors. In some embodiments, the processor 620 can include one or more hardware processors, for example, a microcontroller, a microprocessor, a reduced instruction set computer (RISC), an application-specific integrated circuit (ASIC), an application-specific instruction set processor (ASIP), a central processing unit (CPU), a graphics processing unit (GPU), a physical processing unit (PPU), a microcontroller unit, a digital signal processor (DSP), a field programmable gate array (FPGA), an advanced RISC machine (ARM), a programmable logic device (PLD), any circuit or processor that can perform one or more functions, or any combination thereof.

[0049] To describe a problem, only one processor 620 is shown in the electronic device 600 in the accompanying drawings. However, it should be noted that, the electronic device 600 in this specification can further include a plurality of processors. Therefore, the operations and/or method steps disclosed in this specification can be performed by one processor, or can be jointly performed by a plurality of processors. For example, if the processor 620 in the electronic device 600 performs step A and step B, it should be understood as that step A and step B can be jointly or separately performed by two different processors 620 (for example, a first processor performs step A, and a second processor performs step B, or a first processor and a second processor jointly perform step A and step B).

[0050] FIG. 3 is a flowchart of an information retrieval method P100 according to an embodiment of this specification. An electronic device 600 is deployed with a retrieval platform, and can perform the information retrieval method P100. Specifically, a processor in the electronic device 600 can read an instruction set stored in a local storage medium of the electronic device 600, and then perform the information retrieval method P100 based on a stipulation of the instruction set.

[0051] As shown in FIG. 3, the information retrieval method P100 can include the following steps.

[0052] S110: Receive a retrieval request for requesting to perform information retrieval in a document library, where the retrieval request includes retrieval description information. The document library includes at least one document, and each document includes a level title of at least one level. Text content corresponding to at least some level titles includes at least one

document block, and at least some document blocks include at least one keyword.

[0053] The document library can be understood as a database or a knowledge base of the retrieval platform. In a process in which a user performs information retrieval on the retrieval platform, the retrieval platform performs an information search or information generation based on the document library. The at least one document in the document library can exist in a plurality of forms, and several of these forms are now used as examples.

[0054] Example 1: The document library includes the original at least one document. The original at least one document is shown in FIG. 4 and FIG. 5. A document shown in FIG. 4 is taken as an example. The document includes level titles of a plurality of levels, a level title of a first level is a document name "Platform A", a level title of a second level is "Service provided by the platform A", a level title of a third level is "Individual" and "Merchant", and level titles of a fourth level are "After-sales service" and "Withdrawal service".

[0055] Example 2: The document library includes the original at least one document and M document blocks obtained by dividing the at least one document, where M is an integer greater than or equal to 1.

[0056] Example 3: The document library includes M document blocks obtained by dividing the at least one document.

[0057] In this specification, the document block can be a document segment or a document unit obtained after the document is divided in a preset manner. For example, division can be performed based on a natural paragraph in the document, and one document block corresponds to one paragraph in the document. For example, division can alternatively be performed based on a preset quantity of paragraphs. For example, several adjacent paragraphs are used as one document block. Division into the document block can be performed in a plurality of manners. Examples are not listed one by one in this specification. A person skilled in the art can understand that text content in the same document block is consecutive regardless of a manner of division.

[0058] Each level title of the document illustrated in FIG. 4 and FIG. 5 includes text content (for example, a paragraph A1, a paragraph A2, a paragraph B1, a paragraph B2, ...), but this embodiment of this specification is not limited thereto. Some level titles can include no text content. If the level title includes text content, the text content can be divided into one or more document blocks. For example, text content of the level title "After-sales service" includes paragraphs A4, A5, and A6, and text content of "After-sales service" can be divided into one or more document

blocks.

[0059] The keyword can be understood as an entity in the document block. The entity appears relatively frequently in one or more document blocks, or the entity has a representative meaning, or the entity can be used to locate a document block, etc. The entity includes but is not limited to names such as a person name, a company name, and a place name, a proper noun, etc. For example, the text content of the level title "After-sales service" can be divided into three document blocks. A document block 1 corresponds to the paragraph A4, and mainly describes an operation procedure of refund and return. A document block 2 corresponds to the paragraph A5, and mainly describes an operation procedure of return only. A document block 3 corresponds to the paragraph A6, and mainly describes an operation procedure of refund only. In other words, the document block 1 includes keywords such as "refund and return" and "operation procedure". The document block 2 includes keywords such as "return only" and "operation procedure". The document block 3 includes keywords such as "refund only" and "operation procedure".

[0060] The electronic device 600 can receive retrieval description information entered by the user, or receive retrieval description information sent by another service system. After receiving the retrieval description information, the electronic device 600 can generate a retrieval request that includes the retrieval description information. The retrieval request is used to request to perform information retrieval in the document library. The retrieval description information can be understood as information used to describe a retrieval requirement, for example, a question or a retrieval keyword entered by the user. For example, the retrieval description information can be "Refund operation procedure of a platform A" or "What are main services of xx company?".

[0061] S120: Determine, from the document library based on a knowledge map corresponding to the document library, at least one target document block related to the retrieval description information. The knowledge map represents at least a belonging relationship between level titles of different levels, a belonging relationship between the document block and the level title, and a belonging relationship between the keyword and the document block.

[0062] That the knowledge map corresponds to the document library can be understood that the knowledge map is generated based on the at least one document in the document library.

[0063] For example, the knowledge map can be stored in the document library. Before performing S120, the electronic device 600 can obtain the knowledge map from the document library.

[0064] The following first describes an expression form of the knowledge map.

[0065] In some embodiments, the knowledge map includes a plurality of nodes and edges between the plurality of nodes. The plurality of nodes include a title node representing the level title, a keyword node representing the keyword, and a document block node representing a content summary of the document block. Edges between different title nodes represent the belonging relationship between the level titles at the different levels. An edge between the keyword node and the document block node represents the belonging relationship between the keyword and the document block. An edge between the document block node and the title node represents a belonging relationship between the document block and a level title corresponding to the title node.

[0066] In this embodiment, the knowledge map retains document structure information obtained from an original document, for example, a level title of each level. The knowledge map also analyzes a belonging relationship between different titles, between a title and a file block of the title, between a document block and a keyword. In this case, in a subsequent process in which the electronic device 600 performs retrieval based on the knowledge map, text content corresponding to each node can be traced to a source based on a relationship between nodes in the knowledge map, to avoid a problem that similar nodes are confused, and consequently, a subsequently generated retrieval result is not accurate.

[0067] For example, the electronic device 600 can generate the knowledge map shown in FIG. 6 based on the documents shown in FIG. 4 and FIG. 5. An edge between nodes in the knowledge map can be a directed edge, or can be an undirected edge. Alternatively, an edge between some nodes in the knowledge map is an undirected edge, and an edge between some nodes is a directed edge. The directed edge can better help the electronic device 600 understand a specific relationship between nodes in the knowledge map that are not strongly associated in terms of character content, providing a good basis for subsequent accurate and comprehensive retrieval. In FIG. 6, the directed edge is taken as an example, and an arrow of the directed edge in FIG. 6 is merely taken as an example. This is not limited in this embodiment of this specification. In the knowledge map, title nodes representing level titles of a first level are, for example, "Platform A" and "Platform B"; and title nodes representing level titles of a second level are, for example, "Service provided by a platform A" and "Service provided by a platform B". For a title node representing a level title of another level, references can be made to FIG. 6. Examples are not listed one by one.

[0068] Still as shown in FIG. 6, the knowledge map further includes the document block node

that represents the content summary of the document block. For example, a document block node "A1 summary" represents a content summary of a document block A1 (a document block corresponding to the paragraph A1); and a document block node "B1 summary" represents a content summary of a document block B1 (a document block corresponding to the paragraph B1). Another document block node is understood by analogy. Examples are not listed one by one. As shown in FIG. 6, the electronic device 600 divides each paragraph into one document block, but this embodiment of this specification is not limited thereto. It can be understood that specific characters in nodes such as "A1 summary" and "B1 summary" in FIG. 6 are used as references for describing this embodiment of this specification, to distinguish content summaries of different document blocks. An actual character after the content summary of the document block in this embodiment of this specification is not specifically presented in FIG. 6.

[0069] Still as shown in FIG. 6, the knowledge map further includes keyword nodes such as "return and refund", "return only", "refund only", "operation procedure", "xx company", and "list a product".

[0070] Still as shown in FIG. 6, a directed edge between the title node "Platform A" and the title node "Service provided by a platform A" represents that text content of the level title "Service provided by a platform A" belongs to the level title "Platform A". A belonging relationship represented by a directed edge between other title nodes is understood in the same way. A directed edge between the keyword node "Return and refund" and the document block node "A4 Summary" represents that the keyword "Return and refund" belongs to text content corresponding to a document block A4. A belonging relationship of a directed edge between another keyword node and a document block node is understood in the same way. A directed edge between the document block node "A3 summary" and the title node "Individual" represents that the document block A3 belongs to text content of the title "Individual". A belonging relationship represented by a directed edge between another document block node and a title node is understood in the same way. Details are omitted for simplicity.

[0071] In some embodiments, the plurality of nodes in the knowledge map further include a title summary node that represents a content summary of the level title. An edge between the title node and the title summary node represents a summary relationship between the level title and the content summary of the level title.

[0072] In this embodiment, the knowledge map further covers the title summary node

corresponding to the level title, so that the electronic device 600 can perform supplementary retrieval based on the title summary node corresponding to the level title when related content cannot be retrieved based on the level title, so that retrieval is more comprehensive. In addition, if the electronic device 600 needs to summarize text content of some level titles for the retrieval description information, to obtain a retrieval result, the title summary node is equivalent to pre-summarizing the text content of the level title, so that retrieval efficiency can be improved on the whole.

[0073] For example, as shown in FIG. 7, the knowledge map can further include one or more title summary nodes based on FIG. 6. In other words, all or some of title nodes shown in FIG. 6 can be connected to one title summary node. For example, a title summary node "Document A summary" represents a content summary of the document shown in FIG. 4, and a title summary node "Individual service summary" represents a content summary of text content of the level title "Individual". Another title summary node is understood by analogy. An edge between a title node and a title summary node can be a directed edge, or can be an undirected edge; or an edge between some title nodes and title summary nodes corresponding to the some title nodes is an undirected edge; and an edge between some title nodes and title summary nodes corresponding to the some title nodes is a directed edge. In FIG. 7, the directed edge is taken as an example, and an arrow of the directed edge in FIG. 7 is merely taken as an example. This is not limited in this embodiment of this specification.

[0074] It can be understood that specific characters in title summary nodes such as "document A summary" and "individual service summary" in FIG. 7 are used as references for describing this embodiment of this specification, to distinguish title summary nodes corresponding to different title nodes. An actual character after the content summary of the level title in this embodiment of this specification is not specifically presented in FIG. 7.

[0075] The above-mentioned embodiment describes a presentation manner of the knowledge map. The knowledge map can be generated by the electronic device 600 or another device. The following describes a generation process of the knowledge map by using an example in which the electronic device 600 generates the knowledge map.

[0076] As shown in FIG. 8, a generation method P200 for the knowledge map includes the following steps.

[0077] S210: Extract level titles of N levels from the at least one document, where N is an

integer greater than or equal to 1.

[0078] For example, the electronic device 600 extracts the level title "Platform A" of the first level, the level title "Service provided by the platform A" of the second level, the level title "Individual" and "Merchant" of the third level, and the level titles "After-sales service" and "Withdrawal service" of the fourth level from the document shown in FIG. 4.

[0079] The electronic device 600 extracts the level title "Platform B" of the first level, the level title "Service provided by a platform B" of the second level, and level titles "Sales service" and "Withdrawal service" of the third level from the document shown in FIG. 5.

[0080] For example, the electronic device 600 can extract the level titles of the N levels from the at least one document by using a large model 1. For example, the electronic device 600 generates guidance information 1, and inputs the guidance information 1 into the large model 1. The guidance information 1 indicates a manner of extracting the level title. The large model 1 can extract the level titles of the N levels from the at least one document based on the guidance information 1. The guidance information 1 is, for example, prompt1.

[0081] S220: For each level title, if the level title corresponds to text content, divide the text content to obtain the at least one document block, and perform information extraction on each document block to obtain the keyword and a content summary of the document block.

[0082] For example, the document shown in FIG. 4 is taken as an example. The level title "Platform A" corresponds to text content, and the corresponding text content is the paragraph A1. In this case, the electronic device 600 can divide the paragraph A1 into one or more document blocks. If the paragraph A1 is divided into one document block, denoted as a document block A1, the electronic device 600 needs to perform information extraction on the document block A1, to obtain a keyword and a content summary in the document block A1. For example, for the document block A1, the electronic device 600 learns through extraction that the keyword in the document block A1 is "xx company" and the summary in the document block A1 is "A1 summary". As described above, "A1 Summary" here is merely a reference. Actually, a summary of the document block A1 may be another character description. By analogy, the electronic device 600 performs a similar procedure for each level title corresponding to text content, to obtain one or more document blocks, and obtains a keyword and a content summary that correspond to each document block.

[0083] If a specific level title has no corresponding text content (that is, the level title is closely followed by another level title), the electronic device 600 does not perform S220 for the level title.

[0084] In some embodiments, the electronic device 600 can divide the text content of the level title based on a preset text length, to obtain one or more document blocks.

[0085] In this embodiment, the electronic device 600 divides the text content based on the preset text length, so that text lengths that correspond to different document blocks and that are obtained through division can be relatively balanced, to avoid a case in which when text lengths corresponding to different document blocks have a too large difference, a summary of a long document block is improper or a summarized text is insufficient to comprehensively cover a long document block, making it difficult for the electronic device 600 to perform subsequent retrieval.

[0086] The preset text length can be a preset paragraph length, or can be a preset character length.

[0087] For example, the preset text length is the preset paragraph length. For example, if the preset paragraph length is 2, text content of every two paragraphs is classified into one document block.

[0088] For example, the preset text length is the preset character length. For example, if the preset character length is 500, text content of every 500 characters is classified into one document block.

[0089] For example, the electronic device 600 does not cross level titles when classifying the text content. For example, the text content corresponding to the level title "Platform A" is a paragraph, that is, a paragraph A1. If the preset paragraph length is 2, the electronic device 600 classifies the paragraph A1 into one document block instead of classifying the paragraph A1 and a text content paragraph A2 corresponding to the level title "Service provided by a platform A" into one document block together. In this example, text content corresponding to each document block does not cross two or more level titles, so that the text content corresponding to the document block is prevented from being related to a plurality of aspects, and therefore the text content corresponding to the document block is not summarized easily, affecting subsequent retrieval accuracy.

[0090] S230: Generate the knowledge map based on the level title, the keyword, the content summary of the document block, the belonging relationship between the level titles of different levels, a belonging relationship between the content summary of the document block and the level title, and a belonging relationship between the keyword and the content summary of the document block that are extracted in S210 and S220.

[0091] For example, after the electronic device 600 performs S210 and S220, nodes shown in FIG. 6 can be obtained. In addition, in a process of performing S210 and S220, the electronic device 600 can obtain a belonging relationship between all nodes. After obtaining all the nodes and the belonging relationship between all the nodes, the electronic device 600 can generate the knowledge map shown in FIG. 6 based on all the nodes and the belonging relationship between all the nodes.

[0092] For example, the electronic device 600 can generate a plurality of triples based on all the nodes and a belonging relationship between different nodes, and generate the knowledge map shown in FIG. 6 based on the plurality of triples. The two title nodes "Service provided by a platform A" and "Platform A" are used as examples. A triplet that can be generated based on the two title nodes is, for example, (service provided by a platform A, belonging, platform A), and the triplet represents the two title nodes and a belonging relationship between the two title nodes. Triplets generated based on a belonging relationship between another node and a node are sequentially obtained by analogy. Details are omitted for simplicity.

[0093] In some embodiments, before performing S230, the electronic device 600 can further obtain the content summary of the level title through summarization based on the text content corresponding to each level title, and obtain a summary relationship between content summaries of level titles. That is, in this embodiment, S230 can be replaced with the following:

[0094] The electronic device 600 generates the knowledge map based on the level title, the keyword, the content summary of the document block, the content summary of the level title, the belonging relationship between the level titles of the different levels, the belonging relationship between the content summary of the document block and the level title, the belonging relationship between the keyword and the document block, and a summary relationship between the level title and the content summary of the level title.

[0095] In this embodiment, the electronic device 600 further obtains the content summary of the level title, and presents the content summary in the knowledge map, to provide a more comprehensive retrieval basis for subsequent retrieval.

[0096] For example, for the level title "Platform A", the electronic device 600 summarizes the text content corresponding to "Platform A", to obtain the content summary of "Platform A". The text content corresponding to "platform A" is, for example, all content or the paragraph A1 included in the document shown in FIG. 4.

[0097] For the level title "Service provided by a platform A", the electronic device 600 summarizes text content corresponding to "Service provided by a platform A", to obtain a content summary of "Service provided by a platform A". The text content corresponding to "Service provided by a platform A" is, for example, all content between the title "Service provided by a platform A" and a next-level title at the same level as "Service provided by a platform A", or the paragraph A2.

[0098] A summary of text content corresponding to another level title by the electronic device 600 is similar. Details are omitted for simplicity. The electronic device 600 can also obtain a content summary corresponding to each level title by using a large model 2. For example, the electronic device 600 generates guidance information 2, and inputs the guidance information 2 into the large model 2. The guidance information 2 indicates that text content corresponding to a specific level title is summarized. The large model 2 can obtain a content summary of the level title based on a guidance of the guidance information 2. The guidance information 2 is, for example, prompt2. The large model 1 and the large model 2 can be the same model, or can be different models. This is not limited in this embodiment of the specification.

[0099] For example, the electronic device 600 can generate a plurality of triples based on all the nodes, a belonging relationship between nodes, and a summary relationship that are obtained in this embodiment, and generate the knowledge map shown in FIG. 7 based on the plurality of triples. For an example of the triplet, references can be made to the above-mentioned descriptions. Details are omitted for simplicity.

[0100] In the process of generating the knowledge map in the method P200, the electronic device 600 obtains, through extraction or summarization from the original at least one document, key characters that can highly summarize original document content, to reduce a quantity of characters presented by the nodes in the knowledge map, and preserve an original structure in the document, thereby avoiding confusing similar information in a retrieval process.

[0101] The following describes a manner in which the electronic device 600 determines the at least one target document block in the document library. The electronic device 600 can determine the at least one target document block in a plurality of manners. The following provides descriptions by using three of the plurality of manners as an example. This embodiment of this specification is not limited thereto.

[0102] Manner 1

[0103] In some embodiments, the electronic device 600 determines at least one target node in the knowledge map based on a distance between a vector representation of the retrieval description information and a vector representation of each node in the knowledge map; and further, queries a target mapping relationship based on the at least one target node, to obtain at least one target document block. The target mapping relationship represents a mapping relationship between each node in the knowledge map and the document block.

[0104] For example, a mapping relationship 1 is stored in the document library, and the electronic device 600 can obtain the mapping relationship 1 from the document library. The mapping relationship 1 represents a mapping relationship between each node in the knowledge map and a vector representation of each node. The electronic device 600 can obtain the vector representation of each node in the knowledge map based on the mapping relationship 1. The electronic device 600 quantizes the retrieval description information, to obtain a first vector representation, and determines a distance between the first vector representation and the vector representation of each node in the knowledge map. If a distance between the first vector representation and a vector representation of a specific node is less than or equal to a preset threshold, the electronic device 600 determines the node as a target node. Alternatively, the electronic device 600 performs word segmentation on the retrieval description information, to obtain a plurality of description words; vectorizes each description word to obtain a first vector representation; and for each first vector representation, determines a distance between the first vector representation and the vector representation of each node in the knowledge map. If a distance between at least one of first vector representations corresponding to the plurality of description words and a vector representation of a specific node is less than or equal to the preset threshold, the electronic device 600 determines the node as a target node.

[0105] For example, a target mapping relationship can be further stored in the document library, and the electronic device 600 can obtain the target mapping relationship from the document library. The target mapping relationship represents a mapping relationship between each node in the knowledge map and the M document blocks (obtained through division based on the at least one document in the document library). The document block corresponding to the title node can be understood as some or all document blocks of the level title represented by the title node. The document block corresponding to the document block summary node can be understood as a document block summarized to obtain the document block summary node. The document block

corresponding to the keyword node can be understood as a document block including a character in the keyword node. For example, one paragraph in a document is classified into one document block. A document block corresponding to a title node "Platform A" is the document block A1 (document block corresponding to the paragraph A1), or all document blocks included in the document shown in FIG. 4. Document blocks corresponding to the title node "After-sales service" are the document blocks A4 (document block corresponding to the paragraph A4), A5 (document block corresponding to the paragraph A5), and A6 (document block corresponding to the paragraph A6). A document block corresponding to the document block summary node "A2 summary" is the document block A2 (document block corresponding to the paragraph A2). A document block corresponding to the keyword node "Return only" is the document block A5.

[0106] After determining the at least one target node according to the above-mentioned procedure, the electronic device 600 searches the target mapping relationship to obtain a document block corresponding to each target node, so as to obtain the at least one target document block. For example, after obtaining the document block corresponding to each target node, the electronic device 600 can remove a repetition from the document blocks corresponding to all the target nodes, or obtain a union set of the document blocks corresponding to all the target nodes, to obtain the at least one target document block.

[0107] In Manner 1, the electronic device 600 can match the retrieval description information and the node in the knowledge map, and further query the target mapping relationship, to obtain the target document block. Because a quantity of characters included in each node in the knowledge map is relatively small, matching efficiency can be improved.

[0108] Manner 2

[0109] The electronic device 600 determines at least one first node in the knowledge map based on the distance between the vector representation of the retrieval description information and the vector representation of each node in the knowledge map; and determines, in the knowledge map, at least one second node associated with the at least one first node. The electronic device 600 uses the at least one first node and the at least one second node as the at least one target node; and queries a target mapping relationship based on the at least one target node, to obtain the at least one target document block.

[0110] For example, the electronic device 600 can obtain the vector representation of each node in the knowledge map based on the mapping relationship 1. The electronic device 600

quantizes the retrieval description information, to obtain a first vector representation, and determines a distance between the first vector representation and the vector representation of each node in the knowledge map. If a distance between the first vector representation and a vector representation of a specific node is less than or equal to a preset threshold, the electronic device 600 determines the node as a first node. Alternatively, the electronic device 600 performs word segmentation on the retrieval description information, to obtain a plurality of description words; vectorizes each description word to obtain a first vector representation; and for each first vector representation, determines a distance between the first vector representation and the vector representation of each node in the knowledge map. If a distance between at least one of first vector representations corresponding to the plurality of description words and a vector representation of a specific node is less than or equal to the preset threshold, the electronic device 600 determines the node as a first node.

[0111] After determining the at least one first node based on the above-mentioned procedure, the electronic device 600 can determine the at least one second node based on another node that is in the knowledge map and that is directly or indirectly connected to each first node and an association relationship between the another node and the retrieval description information. For example, the retrieval description information is "What are main services of xx company?". The at least one first node determined by the electronic device 600 based on the retrieval description information includes "xx company". The electronic device 600 can determine that nodes such as "Platform A" and "Platform B" directly connected to "xx company" and "Service provided by a platform A" and "Service provided by a platform B" indirectly connected to "xx company" are associated with the retrieval description information, and therefore, can use the nodes such as "Platform A", "Platform B", "Service provided by a platform A", and "Service provided by a platform B" as second nodes.

[0112] For a procedure in which the electronic device 600 determines the target document block after obtaining the at least one first node and the at least one second node, references can be made to related descriptions in Manner 1. Details are omitted for simplicity.

[0113] In Manner 2, the electronic device 600 can match the retrieval description information and the node in the knowledge map, and mine an associated node; and further, can find a more comprehensive target document block, thereby providing a good basis for obtaining an accurate retrieval result subsequently.

[0114] Manner 3

[0115] The electronic device 600 determines at least one target node in the knowledge map based on a distance between a vector representation of the retrieval description information and a vector representation of each node in the knowledge map; and queries a target mapping relationship based on the at least one target node, to obtain at least one first document block. In addition, the electronic device 600 determines at least one second document block from the at least one document based on a distance between the vector representation of the retrieval description information and a vector representation of each of the at least one document. Therefore, the electronic device 600 determines the at least one target document block based on the at least one first document block and the at least one second document block.

[0116] For an implementation in which the electronic device 600 obtains the at least one first document block, references can be made to related descriptions in Manner 1 or Manner 2. Details are omitted for simplicity. The following mainly describes a process in which the electronic device 600 determines the second document block.

[0117] For example, a mapping relationship 2 is stored in the document library, and the electronic device 600 can obtain the mapping relationship 2 from the document library. The mapping relationship 2 represents a mapping relationship between the M document blocks and a vector representation of each document block. The electronic device 600 can obtain a vector representation of each document block based on the mapping relationship 2. The electronic device 600 quantizes the retrieval description information, to obtain a first vector representation, and determines a distance between the first vector representation and the vector representation of each document block. Similar to the above-mentioned descriptions, the retrieval description information can correspond to a plurality of first vector representations.

[0118] Each document block can correspond to a plurality of vector representations. For example, a plurality of document words are obtained after word segmentation is performed on each document block, and a vector representation corresponding to the document words is obtained after each document word is vectorized. For another example, the vector representation of each document block is divided into a plurality of vector representations based on a preset length. The electronic device 600 can determine the distance between each first vector representation and a plurality of vector representations corresponding to a specific document block. If a distance between at least one of the plurality of vector representations and any first vector representation is

less than or equal to the preset threshold, the electronic device 600 determines that the document block is a second document block. The electronic device 600 can obtain the at least one second document block in the above-mentioned manner.

[0119] After obtaining the at least one first document block and the at least one second document block, the electronic device 600 can remove a repetition from the document block, or obtain the at least one target document block by performing an operation such as obtaining a union set.

[0120] In Manner 3, in addition to matching the retrieval description information and the node in the knowledge map to obtain some target document blocks, the electronic device 600 can directly find some other target document blocks by performing matching with the document block. In other words, in a process of determining the target document block, the electronic device 600 considers a plurality of aspects, to avoid missing a found target document block, and provide a good basis for generating an accurate retrieval result subsequently.

[0121] In some embodiments, the electronic device 600 guides, based on first guidance information, a large model 3 to determine, in the document library based on the knowledge map, the at least one target document block associated with the retrieval description information. The document library and the knowledge map are injected into the large model 3.

[0122] For example, the electronic device 600 can inject the document library and the knowledge map into the large model 3 in advance based on the guidance information 3 (for example, prompt3). Subsequently, the electronic device 600 can guide, based on the first guidance information, the large model 3 to complete S120 in the method P100. The first guidance information is, for example, prompt4. The large model 3 can be the same model as the large model 1 and the large model 2, or can be a different model. This is not limited in this embodiment of this specification.

[0123] In this embodiment, the target document block is determined by using a good understanding capability and a good processing capability of the large model 3 that has original knowledge and injected knowledge, to expect to improve efficiency and accuracy of determining the target document block.

[0124] S130: Generate a retrieval result corresponding to the retrieval request based on the at least one target document block.

[0125] For example, the electronic device 600 can use the at least one target document block

as the retrieval result.

[0126] For example, the electronic device 600 can generate the retrieval result by performing a processing operation such as summarization or deletion on the at least one target document block by using the large model.

[0127] For example, the electronic device 600 guides, based on second guidance information, the large model 3 to generate the retrieval result corresponding to the retrieval request based on the at least one target document block.

[0128] In this embodiment, a relatively refined retrieval result that responds to a question of the user in a targeted manner is generated based on the good text understanding capability, the good summarization capability, and the good generation capability of the large model, to improve accuracy of the retrieval result, and improve retrieval experience of the user.

[0129] It can be understood that before the electronic device 600 performs the information retrieval method P100, related content required in an information retrieval process needs to be recorded into the document library. For example, the electronic device 600 pre-processes the at least one document, the M document blocks, the knowledge map, and each mapping relationship that are mentioned above into a format convenient for subsequent retrieval, and stores the format in the document library.

[0130] It can be further understood that when the information retrieval method P100 is applied to the large model, or is applied to a retrieval platform that can invoke the large model, the large model or the retrieval platform can perform the method P100 based on the retrieval description information entered by the user. Based on the related information of the document block generated based on the document library, the knowledge map, the above-mentioned mapping relationship, etc. can be used as the knowledge base of the large model, to enhance an answer capability of the large model for the question of the user.

[0131] In FIG. 9, that the information retrieval method P100 is applied to the large model is taken as an example. The large model can perform the retrieval process (920). Before performing 920, the operator can inject knowledge into the large model (910). For specific content involved in FIG. 9, references can be made to the above-mentioned descriptions. Details are omitted for simplicity.

It should be noted that, in this embodiment of this specification, a relationship between nodes in the knowledge map is a belonging relationship and a summary relationship is taken as an

example. However, this embodiment of this specification is not limited thereto. For example, the relationship between the nodes in the knowledge map can alternatively be another association relationship that represents an association between nodes. The association relationship includes but is not limited to a parallel relationship, a progressive relationship, or a merge relationship.

[0132] In conclusion, according to the information retrieval method, the device, and the storage medium that are provided in this specification, the target document block related to the retrieval description information can be searched for based on the knowledge map, to generate the retrieval result based on the target document block. In a process of determining the target document block, because the nodes in the knowledge map are relatively simple and refined, a quantity of characters corresponding to the nodes are relatively small. Therefore, efficiency of searching for the target document block can be improved based on the knowledge map. In addition, the knowledge map further represents a level title extracted from an original document, a document block, and a belonging relationship between keywords. Therefore, when different documents include similar content, in the solution provided in this embodiment of the specification, a title, etc. of the similar content can be determined based on the knowledge map, to avoid confusion between the similar content and causing an inaccurate retrieval result. That is, the information retrieval solution provided in this specification can improve retrieval efficiency and accuracy.

[0133] Another aspect of this specification provides a computer-readable non-transitory storage medium, storing at least one group of executable instructions used to perform information retrieval. When the executable instructions are executed by a processor, the executable instructions instruct the processor to implement the steps of the method P100 and/or P200 described in this specification. In some possible implementations, aspects of this specification can further be implemented in a form of a program product, including program code. When the program product runs on an electronic device 600, the program code is used to enable the electronic device 600 to perform the steps of the method P100 and/or P200 described in this specification. The program product configured to implement the above-mentioned method can be a portable compact disc read-only memory (CD-ROM) that includes program code, and can run on the electronic device 600. However, the program product in this specification is not limited thereto. In this specification, a readable storage medium can be any tangible medium that includes or stores a program, and the program can be used by or in combination with an instruction execution system. The program product can be any combination of one or more readable media. The readable medium can be a

readable signal medium or a readable storage medium. The readable storage medium can be, for example, an electrical system, apparatus, or device, a magnetic system, apparatus, or device, an optical system, apparatus, or device, an electromagnetic system, apparatus, or device, an infrared system, apparatus, or device, or a semiconductor system, apparatus, or device, or any combination thereof. More specific examples of the readable storage medium include: an electrical connection including one or more conducting wires, a portable disk, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage component, a magnetic storage component, or any proper combination thereof. The computer-readable storage medium can include a data signal propagated in a baseband or as part of a carrier, and carries readable program code. The propagated data signal can be in a plurality of forms, and includes but is not limited to an electromagnetic signal, an optical signal, or any proper combination thereof. The readable storage medium can further be any readable medium other than the readable storage medium, and the readable medium can be used to send, propagate, or transmit a program used by or in combination with an instruction execution system, apparatus, or component. The program code included in the readable storage medium can be transmitted in any proper medium, including but not limited to a wireless medium, a wired medium, optical cable, an RF medium, or any proper combination thereof. Program code for performing an operation of this specification can be written in any combination of one or more program design languages. The program design language includes an object-oriented program design language, for example, Java or C++, and further includes a conventional procedural program design language, for example, "C" language, or a similar program design language. The program code can be completely executed on the electronic device 600, partially executed on the electronic device 600, executed as an independent software package, partially executed on the electronic device 600 and partially executed on a remote computing device, or completely executed on a remote computing device.

[0134] "Plurality of" mentioned in the embodiments of this specification shall be understood as two or more unless otherwise specified.

[0135] Specific embodiments of this specification are described above. Other embodiments fall within the scope of the appended claims. In some cases, the actions or steps described in the claims can be performed in an order different from that in the embodiments, and the desired results can still be achieved. In addition, the process depicted in the accompanying drawings does not

necessarily need a particular order or consecutive order to achieve the desired results. In some implementations, multitasking and parallel processing are possible or may be advantageous.

[0136] In conclusion, after reading the detailed disclosure, a person skilled in the art can understand that the detailed disclosure can be presented only in an example manner, and can impose no limitation. Although it is not explicitly stated here, a person skilled in the art can understand that the requirements of this specification include various proper changes, improvements, and modifications to the embodiments. These changes, improvements, and modifications are intended to be presented in this specification, and fall within the spirit and scope of the example embodiments of this specification.

[0137] In addition, some terms in this specification are used to describe the embodiments of this specification. For example, "one embodiment", "an embodiment", and/or "some embodiments" mean that particular features, structures, or characteristics described with reference to the embodiments can be included in at least one embodiment of this specification. Therefore, it can be emphasized and understood that two or more of "an embodiment", "one embodiment", or "alternative embodiments" referred in parts of this specification do not necessarily indicate the same embodiment. In addition, particular features, structures, or characteristics can be properly combined in one or more embodiments of this specification.

[0138] It should be understood that in the above-mentioned descriptions of the embodiments of this specification, to help understand a feature, for the purpose of simplifying this specification, various features are combined in a single embodiment, accompanying drawing, or description thereof in this specification. However, this does not mean that the combination of these features is necessary. It is entirely possible for a person skilled in the art to mark some devices and understand the devices as separate embodiments when reading this specification. In other words, the embodiments in this specification can also be understood as an integration of a plurality of secondary embodiments. Content of each secondary embodiment is also true when features are less than all features in one of the above-mentioned disclosed embodiments.

[0139] Each patent, patent application, publication of a patent application, and other materials, for example, articles, books, instructions, publications, documents, and products, that are cited in this specification and that are other than any document that is inconsistent or conflict with this document or that imposes limiting impact on the widest scope of the claims can be incorporated here by reference and used for all purposes now or later associated with this document. In addition,

the term in this document shall prevail when there is any inconsistency or conflict between the description, definition and/or use of a related term in any material and the description, definition and/or use of a related term in this document.

[0140] Finally, it should be understood that the principles of the implementation solutions of this specification are described in the implementation solutions of the application disclosed in this specification. Other modified embodiments also fall within the scope of this specification. Therefore, the embodiments disclosed in this specification are merely examples and not limitations. A person skilled in the art can use an alternative configuration according to the embodiments of this specification to implement the application in this specification. Therefore, the embodiments of this specification are not limited to the embodiments accurately described in the application.

CLAIMS

What is claimed is:

1. An information retrieval method, comprising:

receiving a retrieval request for requesting to perform information retrieval in a document library, wherein the retrieval request comprises retrieval description information, the document library comprises at least one document, each document comprises a level title of at least one level, text content corresponding to at least some level titles comprises at least one document block, and at least some document blocks comprise at least one keyword;

determining, in the document library based on a knowledge map corresponding to the document library, at least one target document block associated with the retrieval description information, wherein the knowledge map represents at least a belonging relationship between level titles of different levels, a belonging relationship between the document block and the level title, and a belonging relationship between the keyword and the document block; and

generating a retrieval result corresponding to the retrieval request based on the at least one target document block.

2. The method according to claim 1, wherein the knowledge map comprises a plurality of nodes and edges between the plurality of nodes, the plurality of nodes comprise a title node representing the level title, a keyword node representing the keyword, and a document block node representing a content summary of the document block;

edges between different title nodes represent the belonging relationship between the level titles at the different levels;

an edge between the keyword node and the document block node represents the belonging relationship between the keyword and the document block; and

an edge between the document block node and the title node represents the belonging relationship between the document block and the level title.

3. The method according to claim 2, wherein the plurality of nodes further comprise a title summary node that represents a content summary of the level title; and

an edge between the title node and the title summary node represents a summary relationship between the level title and the content summary of the level title.

4. The method according to claim 3, wherein the determining, in the document library based

on a knowledge map, at least one target document block associated with the retrieval description information comprises:

determining at least one target node in the knowledge map based on a distance between a vector representation of the retrieval description information and a vector representation of each node in the knowledge map; and

querying a target mapping relationship based on the at least one target node, to obtain the at least one target document block, wherein the target mapping relationship represents a mapping relationship between each node in the knowledge map and the document block.

5. The method according to claim 4, wherein the determining at least one target node in the knowledge map based on a distance between a vector representation of the retrieval description information and a vector representation of each node in the knowledge map comprises:

determining at least one first node in the knowledge map based on the distance between the vector representation of the retrieval description information and the vector representation of each node in the knowledge map;

determining, in the knowledge map, at least one second node associated with the at least one first node; and

using the at least one first node and the at least one second node as the at least one target node.

6. The method according to claim 3, wherein the determining, in the document library based on a knowledge map, at least one target document block associated with the retrieval description information comprises:

determining at least one target node in the knowledge map based on a distance between a vector representation of the retrieval description information and a vector representation of each node in the knowledge map;

querying a target mapping relationship based on the at least one target node, to obtain at least one first document block, wherein the target mapping relationship represents a mapping relationship between each node in the knowledge map and the document block;

determining at least one second document block from the at least one document based on a distance between the vector representation of the retrieval description information and a vector representation of each of the at least one document; and

determining the at least one target document block based on the at least one first document block and the at least one second document block.

7. The method according to claim 1, wherein the determining, in the document library based on a knowledge map corresponding to the document library, at least one target document block associated with the retrieval description information comprises:

guiding, based on first guidance information, a large model to determine, in the document library based on the knowledge map, the at least one target document block associated with the retrieval description information, wherein the document library and the knowledge map are injected into the large model; and

the generating a retrieval result corresponding to the retrieval request based on the at least one target document block comprises:

guiding, based on second guidance information, the large model to generate the retrieval result corresponding to the retrieval request based on the at least one target document block.

8. The method according to claim 1, wherein a generation process of the knowledge map comprises:

extracting the level title of the at least one level from the at least one document;

for each level title, if the level title corresponds to text content, dividing the text content to obtain the at least one document block, and performing information extraction on each document block to obtain the keyword and a content summary of the document block; and

generating the knowledge map based on the level title, the keyword, the content summary of the document block, the belonging relationship between the level titles of different levels, a belonging relationship between the content summary of the document block and the level title, and a belonging relationship between the keyword and the content summary of the document block.

9. The method according to claim 8, wherein the generating process of the knowledge map further comprises:

performing summarization based on text content corresponding to each level title, to obtain a content summary of the level title; and

the generating the knowledge map based on the level title, the keyword, the content summary of the document block, the belonging relationship between the level titles of different levels, a belonging relationship between the content summary of the document block and the level title, and a belonging relationship between the keyword and the content summary of the document block comprises:

generating the knowledge map based on the level title, the keyword, the content summary of

the document block, the content summary of the level title, the belonging relationship between the level titles of the different levels, the belonging relationship between the content summary of the document block and the level title, the belonging relationship between the keyword and the document block, and a summary relationship between the level title and the content summary of the level title.

10. The method according to claim 8, wherein the dividing the text content to obtain the at least one document block comprises:

dividing the text content based on a preset text length, to obtain the at least one document block.

11. An electronic device, comprising:

at least one storage medium, storing at least one instruction set, to perform information retrieval; and

at least one processor, communicatively connected to the at least one storage medium, wherein when running, the at least one processor reads the at least one instruction set, and performs the following operations based on an indication of the at least one instruction set:

receiving a retrieval request for requesting to perform information retrieval in a document library, wherein the retrieval request comprises retrieval description information, the document library comprises at least one document, each document comprises a level title of at least one level, text content corresponding to at least some level titles comprises at least one document block, and at least some document blocks comprise at least one keyword;

determining, in the document library based on a knowledge map corresponding to the document library, at least one target document block associated with the retrieval description information, wherein the knowledge map represents at least a belonging relationship between level titles of different levels, a belonging relationship between the document block and the level title, and a belonging relationship between the keyword and the document block; and

generating a retrieval result corresponding to the retrieval request based on the at least one target document block.

12. The electronic device according to claim 11, wherein the knowledge map comprises a plurality of nodes and edges between the plurality of nodes, the plurality of nodes comprise a title node representing the level title, a keyword node representing the keyword, and a document block node representing a content summary of the document block;

edges between different title nodes represent the belonging relationship between the level titles at the different levels;

an edge between the keyword node and the document block node represents the belonging relationship between the keyword and the document block; and

an edge between the document block node and the title node represents the belonging relationship between the document block and the level title.

13. The electronic device according to claim 12, wherein the plurality of nodes further comprise a title summary node that represents a content summary of the level title; and

an edge between the title node and the title summary node represents a summary relationship between the level title and the content summary of the level title.

14. The electronic device according to claim 13, wherein to determine the at least one target document block, the at least one processor performs the following operations:

determining at least one target node in the knowledge map based on a distance between a vector representation of the retrieval description information and a vector representation of each node in the knowledge map; and

querying a target mapping relationship based on the at least one target node, to obtain the at least one target document block, wherein the target mapping relationship represents a mapping relationship between each node in the knowledge map and the document block.

15. The electronic device according to claim 14, wherein to determine the at least one target node, the at least one processor performs the following operations:

determining at least one first node in the knowledge map based on the distance between the vector representation of the retrieval description information and the vector representation of each node in the knowledge map;

determining, in the knowledge map, at least one second node associated with the at least one first node; and

using the at least one first node and the at least one second node as the at least one target node.

16. The electronic device according to claim 13, wherein to determine the at least one target document block, the at least one processor performs the following operations:

determining at least one target node in the knowledge map based on a distance between a vector representation of the retrieval description information and a vector representation of each node in the knowledge map;

querying a target mapping relationship based on the at least one target node, to obtain at least one first document block, wherein the target mapping relationship represents a mapping relationship between each node in the knowledge map and the document block;

determining at least one second document block from the at least one document based on a distance between the vector representation of the retrieval description information and a vector representation of each of the at least one document; and

determining the at least one target document block based on the at least one first document block and the at least one second document block.

17. The electronic device according to claim 11, wherein to determine the at least one target document block, the at least one processor performs the following operation:

guiding, based on first guidance information, a large model to determine, in the document library based on the knowledge map, the at least one target document block associated with the retrieval description information, wherein the document library and the knowledge map are injected into the large model; and

to generate the retrieval result, the at least one processor performs the following operation:

guiding, based on second guidance information, the large model to generate the retrieval result corresponding to the retrieval request based on the at least one target document block.

18. The electronic device according to claim 11, wherein to generate the knowledge map, the at least one processor performs the following operations:

extracting at least one level title from the at least one document;

for each level title, if the level title corresponds to text content, dividing the text content to obtain the at least one document block, and performing information extraction on each document block to obtain the keyword and a content summary of the document block; and

generating the knowledge map based on the level title, the keyword, the content summary of the document block, the belonging relationship between the level titles of different levels, a belonging relationship between the content summary of the document block and the level title, and a belonging relationship between the keyword and the content summary of the document block.

19. The electronic device according to claim 18, wherein to obtain the at least one document block, the at least one processor further performs the following operation:

dividing the text content based on a preset text length, to obtain the at least one document block.

20. A computer-readable non-transitory storage medium, wherein the computer-readable non-transitory storage medium stores at least one instruction set, and when the at least one instruction set is executed by at least one processor, the method according to any one of claims 1 to 10 is implemented.