ISSN: 2710-0588 DOI: 10.29556/FSE.202203_2(3).0002

Design of Medical Question Answering System based on Knowledge Graph

Qian Tang

School of Dalian Polytechnic University, Dalian 116034, China

Abstract

This paper indroduce the construction of a medical question answering system based on knowledge graph in detail, including data collection, system design and specific functions, and it is pointed out that the application of the system can provide assistance for clinical diagnosis and treatment, and can alleviate the problem of information asymmetry between doctors and patients to a certain extent.

Keywords

Medical Question Answering System; Natural Language Processing; Knowledge Gragh.

1. Introduction

The rapid development of the Internet medical industry provides more convenience for the direct access to medical knowledge through the Internet[1], more and more people choose to consult medical-related issues through various online platforms, but at present, most of these platforms use online question-and-answer exchanges between doctors and patients, so that patients can obtain high-quality medical care. At the same time, the doctor's manual reply is usually expensive and cannot guarantee the actual effect, and most of the questions raised by the patients are very similar. It is necessary to have a way to select the correct answer from a record of the doctor's response to the patient's answer.

The research goal of this paper is an intelligent question answering system for the medical field, providing users with fast and reliable disease-related information and advice. This system collects data from documents and archives, and further builds a knowledge graph after screening, processing and extraction. Functionally, it can be divided into question analysis module, information retrieval module, and question and answer generation module.

2. Demand Analysis

2.1. Problem Analysis Module

People ask questions in various forms, so the medical question answering system needs to parse the questions input by users, and extract keywords of diseases and symptoms from the information input by users.

The problem analysis module is one of the key modules of this system. The responsibility of this module is to extract disease and symptom keywords from the user input information, and identify the user's query intent type. For example, if the user inputs the question "what to do with stomach pain", the module should extract the symptom entity word "stomach pain", mark the word as a symptom, and identify the query intent as query treatment method. Finally, the module transfers the identified entity information and intent type to the information retrieval module for further processing.

ISSN: 2710-0588 DOI: 10.29556/FSE.202203_2(3).0002

2.2. Information Retrieval Module

The computer converts the understood word order into Neo4j query statements, and performs queries in the knowledge base. Since the key-value query in the query language is an exact match, and the user input is often a fuzzy query, the answer retrieval work is mainly completed by calculating the similarity between the query words and words. The relevant entities and attributes are retrieved, and more inputs are given for the selection of relevant entities or attributes, which further enhances the reliability of the answers.

2.3. Answer Generation Module

This module is the main part of the medical question answering system. Further communication with the user, through interaction to get a more accurate case, and finally give the best answer. General users can revise and supplement the questions after the system gives preliminary answers, and finally obtain the best answer through repeated interactions and progressive questions.

3. System Architecture

The system communicates with the patient in the form of natural language dialogue, the patient enters the description of the condition in the form of text, and the system gives the possible disease and its probability and related information about the disease, such as common symptoms, treatment methods, medications, etc. The overall flow of the system is shown in Figure 1.

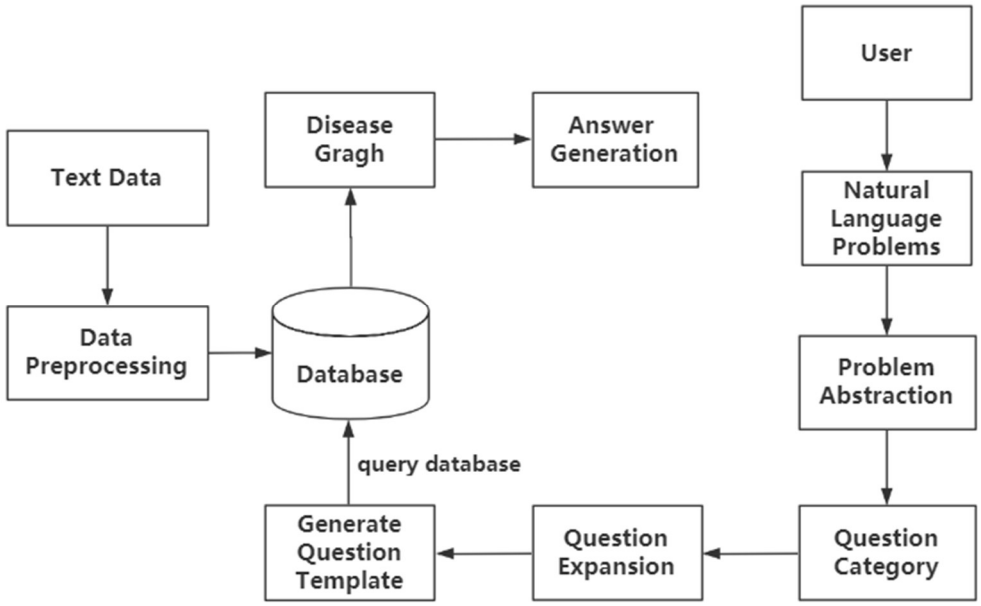

Figure 1. System Architecture Diagram

3.1. Data Acquisition and Processing

The scope of data collection mainly includes literature and archives. The literature data was obtained according to the requirements in the China Traditional Chinese Medicine Database, CNKI, WanFang medical database, about 15,000 documents were obtained, and then the secondary search, literature download, screening were carried out, and finally the literature that met the requirements was selected and the case records were used as the data basis for this study. The final data is further classified and textual extraction is performed on the classified data using text transformation technology.

ISSN: 2710-0588 DOI: 10.29556/FSE.202203_2(3).0002

3.2. Knowledge Graph Construction

Most of the data used in this system is unstructured data, and the automated knowledge extraction methods for such Chinese data are fewer, and the data utilization is more difficult, so this paper adopts knowledge extraction, knowledge fusion and graph database management methods for knowledge graph construction.

Knowledge extraction is mainly to realize the extraction and integration of entity information such as knowledge concepts, attributes, relationships and knowledge classification. Knowledge extraction technology is an indispensable component of a variety of natural language processing technologies such as information extraction, information retrieval, machine translation, and question answering systems[2]. Based on the medical language system, this paper combines Chinese the key technologies of natural language processing such as word segmentation, naming entity extraction[3], and relationship extraction to extract knowledge entity information from medical data, providing data support for the construction of knowledge graph in the medical field.

Knowledge extraction implementation obtains entities, attributes, relationships, and other information from unstructured and semi-structured data, but this information may contain a lot of redundant and erroneous information, and the relationship between data is flattened, lacks hierarchy and logic, and needs to be cleaned and integrated. Knowledge fusion includes two parts: entity linking and knowledge consolidation, through which concept ambiguity can be eliminated, redundant and erroneous concepts can be eliminated, thereby ensuring the quality of knowledge.

Based on the medical language system, the knowledge entity information generated by the knowledge extraction module is used to form a medical knowledge graph and effectively store and manage it. Graph database is a new type of database system that can efficiently handle complex relationship networks based on graph theory algorithms in mathematics[4]. It is good at handling large, complex, interconnected, and changeable mesh data, and its efficiency greatly exceeds that of traditional relational databases.

This article uses the Neo4j graph database to store data. The knowledge graph is usually a network of heterogeneous nodes and edges with semantic relationships, where the nodes represent various knowledge entities, including basic concepts of the field, technical terms, proper nouns, etc[5]., knowledge entities can have multiple attribute information, and the edges represent a variety of relationships between knowledge entities, and the nodes and the edges that connect the nodes together form a network of knowledge graphs[6].

3.3. Knowledge Graph Construction

Through the analysis of data content and manual correction, a knowledge graph of a single document is obtained. The knowledge graph generated by each document is manually labeled and corrected and uploaded to the system as the basic resource for automatic question answering. The system provides a knowledge map of each document to identify the knowledge content and structure of the document, and can also display the map of all documents in the whole system, and can click on each point to view specific information.

3.4. Q&A Implementation

After the user loses the description of the patient's condition, the system first bursts into problem understanding, based on the concept of the medical language system, using the Chinese word segmentation, entity naming and other technologies to transform the natural language asked by the user into a normative knowledge expression; then based on the knowledge spectrum of traditional Chinese medicine to find the relevant knowledge system, judge the disease that the user may suffer from and its probability, give the relevant information of the problem and form a more credible conclusion through the feedback with the user, and at

ISSN: 2710-0588 DOI: 10.29556/FSE.202203_2(3).0002

the same time use the information search function to search for cases with a certain degree of similarity[7]. Export prescription recommendations and provide users with adjunctive diagnosis and treatment recommendations.

4. Conclusion

This paper introduces an automated question-and-answer system based on key technologies such as literature collection, knowledge graph construction, information retrieval, and medical knowledge feedback. The system draws knowledge entities such as symptoms and medication from electronic medical records, medical cases, medical guidelines, and medical literature to construct a knowledge graph. According to the questions and auxiliary information raised by the user, the system gives answers through knowledge graph search and intelligent question-answer technology, including the diseases that patients may have, reference cases and their probabilities, etc., to provide assistance for clinical diagnosis and treatment. The system can carry out intelligent search, through the retrieval to provide users with valuable data information, users can efficiently, accurately and quickly query relevant information, help users self-diagnose, to a certain extent to alleviate the problem of doctor-patient information asymmetry.

References

[1] Lu S, Xu X, Chen Y, et al. Research on Medical Knowledge Question Answering System Based on Artificial Intelligence Technology. Software Guide. 2019, 18 (12): 65-68.

[2] Tang Y, Wang L, Chen X, et al. Knowledge Extraction: Automatic Classification of Matching Rules[M]. 2021.

[3] Nesterov A, D Umerenkov. Distantly supervised end-to-end medical entity extraction from electronic health records with human-level quality[J]. 2022.

[4] Jia L, Liu L, Liu J, et al. Design and Construction of Knowledge Question Answering System Based on Traditional Chinese Medicine Language System [J]. Chinese Journal of Medical Library and Information Science, 2019, 28 (5): 11-14.

[5] Paulheim, Heiko. Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. 1 Jan. 2017 : 489 – 508.

[6] Rotmensch M , Halpern Y , Tlimat A , et al. Learning a Health Knowledge Graph from Electronic Medical Records[J]. Scientific Reports, 2017, 7(1).

[7] Liang H, Lin K, Zhu S. Short Text Similarity Hybrid Algorithm for a Chinese Medical Intelligent Question Answering System[J]. Springer, Singapore, 2019.