applied sciences

MDPI

Article

Research on Hierarchical Knowledge Graphs of Data, Information, and Knowledge Based on Multiple Data Sources

Menglong Li [1,*], Zehao Ni [2], Le Tian [1], Yuxiang Hu [1], Juan Shen [1] and Yu Wang [1]

[1] Institute of Information Technology, PLA Strategic Support Force Information Engineering University, Zhengzhou 450002, China
[2] Institute of Artificial Intelligence, Zhengzhou Railway Vocational & Technical College, Zhengzhou 450002, China
* Correspondence: growinglml@163.com

Abstract: In the existing medical knowledge graphs, there are problems concerning inadequate knowledge discovery strategies and the use of single sources of medical data. Therefore, this paper proposed a research method for multi-data-source medical knowledge graphs based on the data, information, knowledge, and wisdom (DIKW) system to address these issues. Firstly, a reliable data source selection strategy was used to assign priorities to the data sources. Secondly, a two-step data fusion strategy was developed to effectively fuse the processed medical data, which is conducive to improving the quality of medical knowledge graphs. The proposed research method is for the design of a multi-data-source medical knowledge graph based on the DIKW system. The method was used to design a set of DIK three-layer knowledge graph architectures according to the DIKW system in line with the medical knowledge discovery strategy, employing a scientific method for expanding and updating knowledge at each level of the knowledge graph. Finally, question and answer experiments were used to compare the two different ways of constructing knowledge graphs, validating the effectiveness of the two-step data fusion strategy and the DIK three-layer knowledge graph.

Keywords: medical knowledge graph; DIKW; multi-data-source; question and answer

Citation: Li, M.; Ni, Z.; Tian, L.; Hu, Y.; Shen, J.; Wang, Y. Research on Hierarchical Knowledge Graphs of Data, Information, and Knowledge Based on Multiple Data Sources. *Appl. Sci.* **2023**, *13*, 4783. https://doi.org/10.3390/app13084783

Academic Editor: Malik Yousef

Received: 12 February 2023
Revised: 16 March 2023
Accepted: 24 March 2023
Published: 11 April 2023

1. Introduction

The development of information technology continues to promote the transformation of Internet technology, and Web technology, as the iconic technology of the Internet era, is at the core of this transformation. Human beings have experienced the Web 1.0 era, characterized by document interconnection, and the Web 2.0 era, characterized by data interconnection, and are moving towards a new Web 3.0 era based on knowledge interconnection [1]. The purpose of knowledge interconnection is to build a World Wide Web that both humans and computers can understand, making the network more intelligent. In order to represent the whole interconnected human cognitive world more profoundly, a new method of knowledge representation and management has been developed: the knowledge graph. Knowledge graphs are the foundation and bridge for the realization of intelligent semantic retrieval, laying the foundation for knowledge interconnection in the World Wide Web.

The concept of the knowledge graph was formally proposed by Google in May 2012 and its development continued after 2013 with the continuous progress in intelligent information services and applications. Today, knowledge graphs are not only a research hotspot in academia but also an indispensable application technology for intelligent services in industry.

Medicine is one of the vertical fields where knowledge graphs are most widely used, and it is also a hotspot of research for the field of artificial intelligence in China and abroad, including research in disease risk assessment, intelligent assisted diagnosis and treatment,

medical quality control, and medical knowledge questions and answers. In addition, many companies have built their own knowledge graphs, such as the Yi Zhi Lu medical think tank at Ali Health (https://www.mdeer.com, accessed on 7 December 2022), the AI medical knowledge graph Artificial Intelligence + Professional + GC (APGC) developed by Sougo, and other applications. In the medical field, typical medical knowledge maps include the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT) (https://www.snomed.org, accessed on 8 January 2023), Watson Health (https://www.ibm.com/watson, accessed on 8 January 2023) developed by International Business Machines (IBM) Corporation, and the Chinese medicine knowledge map developed by Shanghai Shuguang Hospital in China.

With the ongoing development of medical informatization, a huge amount of medical data has been generated. In the face of massive semi-structured and unstructured medical resources, quickly discovering relevant information useful to users is a major challenge. Moreover, the recent research on the construction of medical knowledge graphs has focused on a single source of medical data, which cannot meet the actual needs of users well.

Because of the above issues, we propose a method for the construction of DIK medical knowledge graphs with multiple data sources in this paper. We first used different methods to collect data from different medical data sources, and this part of the work was described in a previous publication [2]. Among the sources used, the BiLSTM-CRF model and Web crawler data extraction were used for Chinese electronic medical records and online medical communities, respectively. Secondly, in the data classification stage, to give full play to the values of the different medical data sources, a data source selection strategy was developed, as described in this paper. This strategy was used to divide the data categories according to the characteristics of different medical data sources in order to make the best use of the medical data. In addition, to improve the effectiveness of data fusion with different data sources, a two-step data fusion strategy is proposed: (a) disease entity standardization; (b) disease entity-centric structuring and acquisition of triple-structured data. Then, the data, information, knowledge, and wisdom (DIKW) system was used as the basis for modeling a DIK three-layer knowledge graph, and it was abstracted into a generic method model so that it can be implemented as a perfect knowledge discovery strategy for massive medical resources. To better adapt to the knowledge iteration update, a set of scientific methods for extending and updating knowledge was developed for each layer of the DIK knowledge graph.

Finally, a preliminary implementation of a question-and-answer system based on the DIK knowledge graph was undertaken in conjunction with Flask and Python, and the differences between the two approaches to knowledge graphs were compared and validated. The question and answer system was capable of answering "what/why/how/who"-type questions in natural language. The complete framework structure is shown in Figure 1.

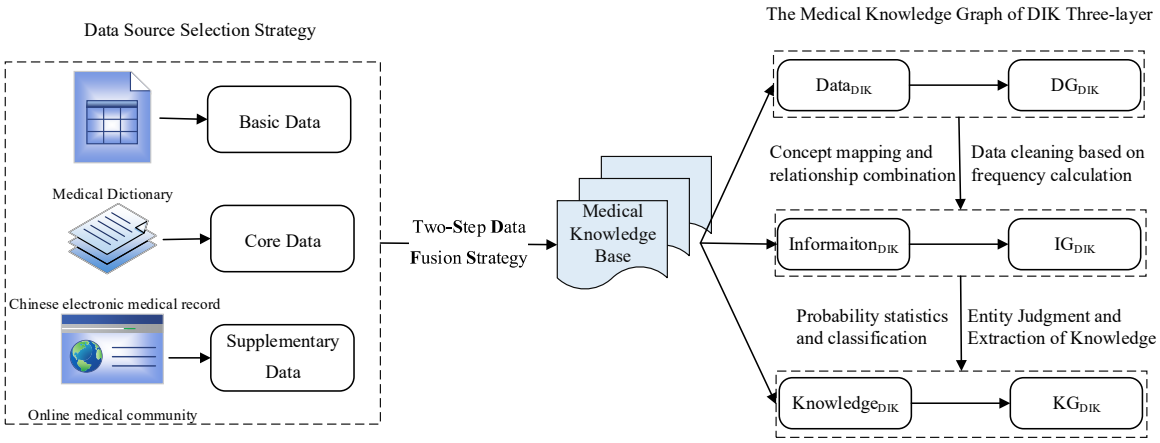

Figure 1. The construction process of a multi-data source medical knowledge graph based on the DIK system.

The innovative points of the proposed method in this paper are as follows:

(a) The characteristics of different medical data sources are used to improve the generalization of medical knowledge graphs;
(b) According to the commonality of medical data, a two-step data fusion strategy was proposed, which is beneficial to improve the effective data fusion between different data sources;
(c) According to the DIKW system, a construction method of the DIK medical knowledge graph applicable to the medical field was proposed.

2. Related Work

In the field of medicine, with the development of regional health informatization and medical information systems, a large amount of medical data has been accumulated. How to extract information from these types of data, and manage and apply them, is a key issue in promoting medical intelligence and is also a medical knowledge retrieval, clinical diagnosis, medical quality management, electronic medical records, and the basis of intelligent processing of health records [3]. Liu [4] designed and implemented a knowledge search system based on a Chinese medical knowledge graph using JAVA, which can process the natural language input by the user such as syntactic analysis and semantic dependency analysis, identify the user's search intent, and return the user's required knowledge in a more intuitive and precise way with the help of the knowledge graph. Zhou [5] developed a knowledge graph-based question-and-answer system to address the problems of traditional search engines and to achieve accurate answers to questions asked by users. The system includes a data collection module, a Q&A module, and a back-end management and front-end display module. Wang et al. [6] explored automated construction methods and standardized processes for TCM (Traditional Chinese Medicine) knowledge graphs using text extraction, relational data transformation, and data fusion to achieve template-based TCM knowledge Q&A and assisted prescribing based on knowledge graph reasoning. However, the abovementioned research on the construction of medical knowledge graphs had no standard data collation methods and a single source of medical data, which fails to meet the diverse needs of different users.

Han et al. [7] proposed a theoretical framework for the construction of medical knowledge graphs with the fusion of multiple data sources, starting from key technologies such as medical big data acquisition, medical entity, and relationship annotation, medical entity recognition, medical entity linking, medical entity relationship mining, and Chinese medical knowledge graph representation and storage. However, they did not integrate any of the knowledge discovery strategies into the medical knowledge graph framework, resulting in the inability to provide guidance to users for knowledge discovery from the massive number of medical resources.

Cowie et al. [8] divided information extraction into three levels: entity, relationship, and attribute, and extended the concept based on the existing knowledge graph, which is further subdivided into data graph, information graph, knowledge graph, and wisdom graph, and can be used to answer questions related to the 5W-related questions [9]. Through the survey, it was found that the current knowledge graph research field can be divided into general knowledge graphs and industry knowledge graphs. The industry knowledge graph needs to consider different business scenarios and users, so the attributes and data patterns of entities are richer. The DIK medical knowledge graph to be constructed in this paper belongs to the industry knowledge graph.

3. Medical Data Sources

3.1. Date Source Classification

Since different categories of medical data contain different medical knowledge, the value of medical big data can be realized only by fusing multiple medical data. For example, Wu et al. [10] proposed the method of fusing multiple data resources to construct knowledge graphs to improve the practical application value of knowledge graphs. Therefore, in this

paper, three medical data resources, i.e., medical dictionaries, electronic medical records, and online medical communities, were selected to improve the practical application value of medical knowledge graphs. To be able to effectively enhance the richness of the medical knowledge graph, this paper counted characteristics of data sources selected as medical data sources and formulated a data source selection strategy for the construction of multi-data source medical knowledge graphs. The statistical results and specific characteristics are described as follows.

(a) Medical Dictionary

The medical dictionary mainly includes existing medical dictionary resources, such as the International Classification of Diseases Manual ICD-11, etc. These resources are highly professional and are among the most important data sources of the medical knowledge graph.

The International Classification of Disease (ICD) is a manual for classifying diseases published by the World Health Organization (WHO). ICD-11 (http://www.medsci.cn/sci/icd-10.asp, accessed on 10 December 2022) is the 11th revision of ICD, and China has participated in the revision, research, and development process of ICD-11. As shown in Figure 2, some examples of disease codes from the ICD-11 disease classification manual are shown. The first column is the disease name, and the second column is the disease code, which is used to uniquely identify a disease entity. For the first time, the ICD-11 has an all-electronic version, which is easy to use and reduces error rates. A total of 55,000 codes were included in this compilation, far more than the 14,400 codes of the ICD-10. At present, the Chinese translation version provided by the State Hospital Administration is the ICD-11 Concise Code List, with a total of 32,198 entries.

Hypertensive encephalopathy	8B22.8	**ICD-11 disease coding**
Hypertensive retinopathy	9B71.1	**ICD-11 disease coding**
Hypertension	L1-BA0	**ICD-11 disease coding**
Essential hypertension	BA00	**ICD-11 disease coding**
Diastolic hypertension was combined with systolic hypertension	BA00.0	**ICD-11 disease coding**
Isolated diastolic hypertension	BA00.1	**ICD-11 disease coding**
Isolated systolic hypertension	BA00.2	**ICD-11 disease coding**
Other special refers to essential Hypertension	BA00.Y	**ICD-11 disease coding**
Essential hypertension, not specified	BA00.Z	**ICD-11 disease coding**
Hypertensive heart disease	BA01	**ICD-11 disease coding**
Hypertensive nephropathy	BA02	**ICD-11 disease coding**
Hypertensive crisis	BA03	**ICD-11 disease coding**
Secondary hypertension	BA04	**ICD-11 disease coding**

Figure 2. The ICD-11 part of the disease coding example.

(b) Chinese electronic medical record

Chinese electronic medical record is a record of clinicians' diagnosis and treatment process of patients, mainly including discharge summary and various treatment records, which contains many professional terms in the medical field and a high density of entity vocabulary and is an important data source for constructing medical knowledge graph.

With the support of local hospitals, we obtained 2000 Chinese electronic medical records, and the statistical results are shown in Table 1. The main entity types in the outpatient records and consultation records are body parts, symptoms, and examinations; the main entity types in the medical history records and discharge records are body parts, symptoms, and examinations.

Table 1. Statistical results of various medical entities in electronic medical records.

Medical Category	Entity				
	Body Parts	Symptom	Examination	Disease	Treatment
Outpatient Clinic	1810	3750	10	720	0
Medical History	39,200	41,230	34,680	3520	400
Treatment	4900	2970	3290	1590	4893
Discharged From Hospital	31,900	21,730	29,400	80	98
Total	77,810	69,680	67,380	5910	5391

(c) Online medical community

The advantages of online medical communities as an emerging source of medical data are shown below:

(a) There are plentiful medical data resources available for mining;
(b) Medical data are originated from the real situation of users;
(c) These data have better timeliness and can update faster.

In addition, online medical data do not contain private data, are publicly available, and there is a large amount of data. Although the data quality is low compared with the first two types of medical data, it is improving day by day and serve as an important complementary source of data for the medical knowledge graph.

We used the web crawler Scrapy (https://github.com/scrapy/scrapy, accessed on 23 May 2022) to crawl medical data from authoritative online medical communities such as Thumb Doctor (http://muzhi.baidu.com, accessed on 23 May 2022), Seeking Medicine, and Doctor Seeker (http://www.xywy.com, accessed on 23 May 2022) and performed statistical analysis on the crawled data, as shown in Table 2.

Table 2. Statistical results of medical entities in the online medical community.

Medical Entity	Disease	Drug	Food	Doctor
Quantity	8807	4828	4870	100

3.2. Data Strategy

3.2.1. Data Source Selection Strategy

The data source selection strategy developed in this paper first took the quality of data sources as a precondition to ensure the accuracy of medical data. Then, according to the statistical results of corresponding data sources as the applicable standard, the data characteristics of different data sources were utilized to the maximum use value. Specific implementation details are described below:

(a) Firstly, considering that the publicly available medical dictionaries possess expertise, they were used as the base data to provide standards for medical knowledge graphs, for example, standards for disease entity names and coding rules;
(b) Secondly, clinical experience data have higher authority and validity, so using electronic medical records as the core data of a medical database facilitates improving the accuracy of intelligent diagnosis, for example, detailed symptom narratives, clinical manifestations, drug effects, and treatment;
(c) Finally, medical community data sources can not only greatly enrich the diversity of medical databases, but also guarantee the timeliness of medical knowledge using the

Web. However, considering the quality of the data, they are used as supplementary data to the medical database, for example, dietary issues, usage, dosage, precautions of drugs, medical science, etc.

Considering the openness and availability of data, we chose publicly available medical dictionaries, online medical communities, and a certain number of electronic medical records obtained from local hospitals as the medical database in this paper. Among them, electronic medical records were desensitized data after manual processing, contain no private data of users, and were only used for academic research.

3.2.2. Two-Step Data Fusion Strategy

Data fusion is a key step in constructing knowledge graphs from multi-data sources. Wu Yunbing et al. first constructed domain ontology libraries and used rules such as similarity detection and conflict resolution to fuse ontology libraries of multiple domains to form a global ontology library. They targeted the construction of multi-data source knowledge graphs for generic domains and gave a standardized process from data acquisition, data processing, data fusion, and knowledge graph construction, but it could not be directly used in the medical domain [10]. By researching the literature in the medical field, we found that disease is the most critical entity, doctors need to diagnose diseases through symptoms, treatments need to be selected for different diseases, tests need to be performed to accurately determine diseases, etc. Other medical entities need to cross each other through disease entities. We concluded that the disease entity has the characteristic of a "transportation hub" among many predefined medical entities. Based on this characteristic, this paper proposed a two-step data fusion strategy for medical data.

In addition, the target of data fusion is the integration of data and knowledge from different data sources [11]. The goal of data fusion methods used in multi-data source knowledge graphs is only the data entities themselves, ignoring the fusion after forming a triad of data in the form of "entity1-relationship type-entity2". In medical data, there are many different types of disease entities, and it is easy to fuse two disease expressions that are similar but have different meanings into one entity for processing. To avoid causing such errors, this paper prioritized the different data sources according to their characteristics in the data source selection strategy. Moreover, in entity alignment, the medical dictionary is used as the entity alignment standard to avoid this wrong operation of fusing different entities into the same concept.

The commonly used data fusion methods are to process medical entities from different data sources acquired by different methods through the same entity alignment method, which fails to effectively resolve entity conflicts. The two-step data fusion strategy proposed in this paper added the type of data sources (basic data, core data, and supplementary data) as a consideration to ensure the correctness of entity alignment operation when entity conflicts are encountered. As a result of this process, the triplet data "entity1-relationship type-entity2" was processed twice. Finally, according to the entity characteristics of " hub" in medical data, medical entities from different data sources were fused according to medical relationship type and data source classification (basic data, core data, supplemental data) based on disease entity names to obtain the final triplet data.

The two-step data fusion strategy first adopted the most authoritative medical dictionary data source as the alignment standard in the entity alignment process, avoiding the wrong operation of fusing similar disease entities into the same entity. Then, according to the entity characteristics of a "hub" in the medical field, the medical data from three data sources were fused according to their respective functions, and the medical entities from different data sources were linked into a "treatment plan" in the form of knowledge through disease entities. With the abovementioned two characteristics, we not only improved the effectiveness of data fusion in terms of entity fusion but also improved the effectiveness of knowledge fusion in terms of triplet fusion.

4. The DIK Architecture

4.1. The Definition of DIK Medical Resources

Based on the expansion of the knowledge graph in the reference [12], this paper proposed a knowledge discovery strategy based on the DIK (Data, Information, and Knowledge) system for constructing a DIK three-layer medical knowledge graph with resource elements and three-layer graphs defined as follows:

Definition 1. *(Resource Elements) The resource element includes three forms: Data Resource ($Data_{DIK}$), Information Resource ($Information_{DIK}$), and Knowledge Resource ($Knowledge_{DIK}$).*

$$Elements_{DIK} = < Data_{DIK}, Information_{DIK}, Knowledge_{DIK} > \quad (1)$$

Definition 2. *(Graphs) Expanding the knowledge graph concept into a three-layer DIK knowledge graph: Data Graph (DG_{DIK}), Information Graph (IG_{DIK}), and Knowledge Graph (KG_{DIK}).*

$$Graph_{DIK} = < DG_{DIK}, IG_{DIK}, KG_{DIK} > \quad (2)$$

According to the construction process of the knowledge graph, we first performed data extraction of semi-structured medical data such as electronic medical records and online medical communities. Secondly, we performed data screening and data cleaning on the extracted data to make the data appear structured and modeled. Thirdly, we performed data integration, statistical analysis, and comprehensive induction to form knowledge. Finally, we performed tacit knowledge mining to provide users with personalized medical services.

The DIK three-layer medical knowledge graph model constructed in this paper was concretely represented by a static analysis centered on entity synthesis calculation on DG_{DIK} to a dynamic automatic abstraction resource optimization process on IG_{DIK} and KG_{DIK}, and supports compatible empirical knowledge introduction and efficient automatic semantic analysis.

The DIKW mechanism is a hierarchy of progressive relationships, mining from $Data_{DIK}$ to $Information_{DIK}$, obtaining $Knowledge_{DIK}$ from $Information_{DIK}$, and finally abstracting $Wisdom_{DIK}$ from $Knowledge_{DIK}$. This paper semantically modeled healthcare resources based on the first three layers of DIKW at the DG_{DIK} layer to calculate two data frequencies of the resource element $Data_{DIK}$, and on IG_{DIK} and KG_{DIK} to analyze the resource element $Information_{DIK}$ and $Knowledge_{DIK}$ respectively for automatic abstraction of the resource optimization process to achieve the introduction of compatible empirical knowledge and automatic semantic analysis, as shown in Table 3.

Table 3. Explanation of medical resource types.

Resource Element	$Data_{DIK}$	$Information_{DIK}$	$Knowledge_{DIK}$
form	Discrete Resource Elements	concept portfolio	Classification and Abstraction
Resource Answers	Who/When	what	why/how
map type	DG_{DIK}	IG_{DIK}	KG_{DIK}
use	Identify resource existence	Interaction and Collaboration	Inference and Prediction

4.2. The Architecture of the DIK Three-Layer Knowledge Graph

This section describes in detail the knowledge discovery principles and the specific implementation process in the DIK three-layer medical knowledge graph.

4.2.1. Data Layer

DG_{DIK} indicates the data layer medical knowledge, which in essence is something calculated as a static analysis based on the entity composite degree. Therefore, the entity

$Data_{DIK}$ in DG_{DIK} is a discrete resource element. DG_{DIK} can record the frequency of occurrence of $Data_{DIK}$ by defining the frequency as a 2-tuple structure: Data Frequency $<f_{structure}, f_{spatial}>$.

Definition 3. *DFreq is a 2-tuple definition of the data frequency of the resource element* $Data_{DIK}$*, as shown in Equation (3).* $f_{structure}$ *is the structural frequency, which indicates the number of times* $Data_{DIK}$ *appears in different data structures. In the medical field, it can be defined as the number of times the disease entity appears in different treatment measures.* $f_{spatial}$ *is the spatial frequency, which indicates the number of times* $Data_{DIK}$ *appears in different spatial locations, and in this paper, we expand it to the number of times the disease appears in different medical departments.*

$$DFreq = < f_{spatial}, f_{structure} > \quad (3)$$

When modeling the resource elements $Data_{DIK}$ in the DG_{DIK} layer, they are discrete resource elements of numbers and other types of information obtained by statistical analysis, which allows them only to be analyzed statically and there to be no dynamic prediction of them. Therefore, they have no meaning in themselves without contextualization, and statistical examples of data frequency are shown in Figure 3.

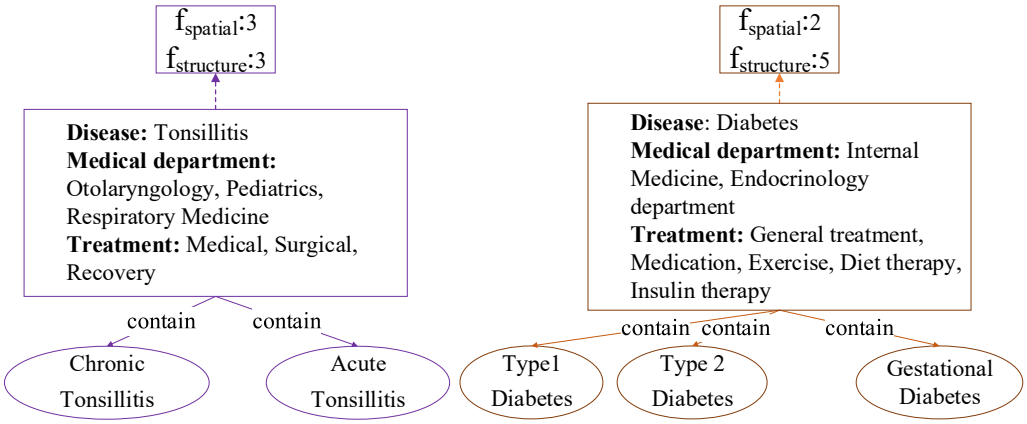

Figure 3. Statistics of $Data_{DIK}$ on $f_{spatial}$ and $f_{structure}$.

4.2.2. Information Layer

The IG_{DIK} represents the information layer knowledge graph and improves the cohesiveness of IG_{DIK} nodes by performing initial abstraction of concept mapping and relationship combination between the interaction degree and context mining of $Data_{DIK}$ entity nodes in DG_{DIK} to obtain the $Information_{DIK}$ node [12]. In DG_{DIK}, only the data frequency was counted, and the accuracy of the acquired data was not analyzed, so a large amount of redundant data was generated. Therefore, in the IG_{DIK} layer, data cleaning was performed to eliminate redundant data, and preliminary abstraction was performed to improve the cohesiveness of the data based on the interaction degree between entity nodes. To be more integrated with the real situation in healthcare and to improve the significance of statistical data, for the recording of the interaction frequency between entity nodes on IG_{DIK}, we only consider the direction of interaction between entities and not the type of interaction. Because different nodes may have the same interaction type or a wide variety of interaction types (reducing statistical significance), but the in-degree and out-degree of each node can be precisely quantified statistically.

In order to reduce redundant data, we defined the composite degree to measure the importance of nodes in IG_{DIK}, and the calculation formula is defined as follows:

$$\text{Com\_degree} = \sqrt{\deg^+ \times \deg^-} \quad (4)$$

As shown in Equation (4), deg^+ and deg^- are the in-degree and out-degree of the node, respectively. However, the composite degree on IG_{DIK} only counts the interaction frequency of nodes, and there is no improvement compared with the frequency statistics in DG_{DIK}. As shown in Figure 4, Entity1 and Entity2 are low-frequency nodes on DG_{DIK}, and Entity3 and Entity4 are high-frequency nodes on DG_{DIK}, i.e., $DFreq = (Entity1, Entity2) < DFreq = (Entity3, Entity4)$. However, the Com_degree on IG_{DIK} is the opposite result, i.e., $Com\_degree = (Entity1, Entity2) > Com\_degree = (Entity3, Entity4)$. Therefore, only using the composite degree is used to measure the nodes in IG_{DIK}, which tends to lose information. To measure the importance of nodes further accurately in IG_{DIK}, we defined the calculation of Impor for retaining the information of nodes on DG_{DIK}. where α and β denote the weight coefficients of the DFreq of entity nodes on DG_{DIK} and Com_degree on IG_{DIK}, respectively, for measuring the importance of nodes, both of which can be obtained by simple learning model training.

$$Impor = \alpha DFreq \times \beta Com\_degree \tag{5}$$

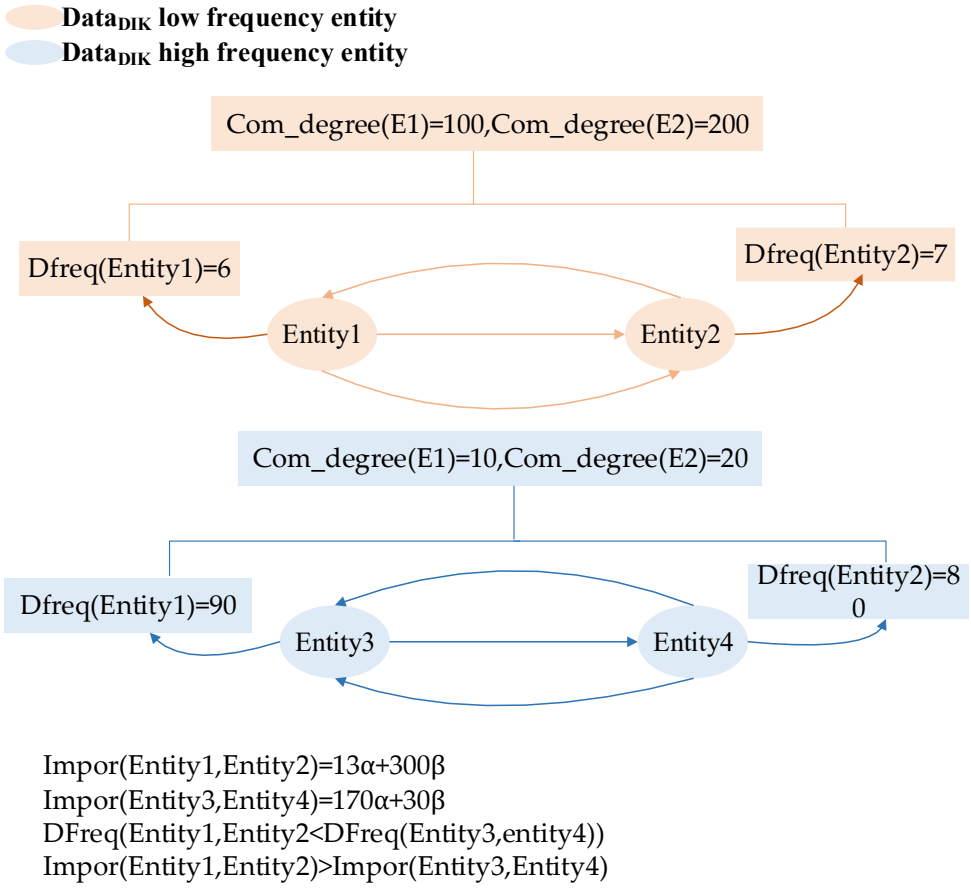

Figure 4. Measuring the importance of entity nodes in $Information_{DIK}$.

By mining the $Data_{DIK}$ elements in DG_{DIK}, a new concept $Information_{DIK}$ was generated in IG_{DIK}, which reflected multiple interactions between nodes. For example, the $Data_{DIK}$ element "pharyngitis, rhinitis, tonsillitis", etc., is integrated to improve the expression of the medical knowledge graph, and the $Information_{DIK}$ element "ENT disease" is obtained. Therefore, we calculated the internal interaction degree and external interaction degree by circling a specific number of entities, as shown in Equation (6).

$$cohesion = \frac{IFreq_{II}}{IFreq_{EI}} \tag{6}$$

Cohesion is the ratio of internal interaction to external interaction and is a measure of the degree of association between entity nodes. $IFreq_{EI}$ is the number of external interactions between entity nodes, and $IFreq_{II}$ is the number of internal interactions between entity nodes. The integration of different entities with maximum cohesion into the same concept is used to enhance the cohesiveness of the model and to improve the abstraction of information. The newly integrated concepts are marked as new nodes on IG_{DIK}, and the structural frequency and spatial frequency of the new nodes were recounted on the DG_{DIK} layer.

4.2.3. Knowledge Layer

The KG_{DIK} is a knowledge layer graph obtained by refining the links between the $Information_{DIK}$ in IG_{DIK} and integrating and summarizing the $Information_{DIK}$ node in IG_{DIK}, integrating and summarizing the InformationDIK to transform it into the $Knowledge_{DIK}$ node. The KG_{DIK} layer contains various semantic relations that perform information reasoning and entity linking. Information reasoning requires the support of relevant relational rules, which are constructed manually and often take a lot of time and effort. The path ordering algorithm used each different relational path as a one-dimensional feature to extract relations by constructing many relational paths in KG_{DIK} to build feature vectors for relational classification and relational classifiers. In this paper, Formulas (7) and (8) were used to measure the accuracy of the extracted relations and the importance of the fully evaluated nodes, respectively, to improve the medical knowledge graph representation. $P(E_1 \rightarrow E_2)$ denotes a path from E_1 to E_2. Q denotes all complete paths between two entities, π denotes a path. $\theta(\pi)$ denotes the weight of the path π, and this relation is considered to be established when the Cr exceeds a preset threshold, and the weight of the path as well as the Cr can be obtained by training. Final_Impor denotes the importance of the fully evaluated node on KG_{DIK}, λ is the weight of the relation Rel, and n is the number of relation types.

$$Cr(E_1, R, E_2) = \frac{\sum_{\pi \in Q} P(E_1 \rightarrow E_2)\theta(\pi)}{|Q|} \tag{7}$$

$$Final\_Impor = Impor \times \gamma \frac{\sum_{i=1}^{n} \lambda_i \times Rel_i}{n} \tag{8}$$

Final_Impor integrated the characteristics of the DIK three-layer graph considering the data frequency DFreq on DG_{DIK}, the interaction degree Impor between entity nodes on $IG_{DIK,}$ and the semantic relationship type on KG_{DIK}. By combing the importance of nodes on the three-layer graph comprehensively and evaluating the importance of nodes, we avoided losing some nodes with low frequency but containing important relational interactions, thus improving the data integrality. Finally, KG_{DIK} was defined as a directed graph, where the relationships between nodes were directed and differentiated, which could satisfy more different kinds of semantic relationships.

4.3. Implementation

The general construction process of medical knowledge graph can be summarized into three modules, namely, medical data extraction, medical knowledge fusion, and medical knowledge inference. Additionally, according to the data to wisdom knowledge discovery in the DIKW system, the DIK medical knowledge graph construction can also be summarized into three modules: statistical $Data_{DIK}$, integration $Information_{DIK}$, and acquisition $Knowledge_{DIK}$. As shown in Figure 5, the construction process of both knowledge maps is a data $\rightarrow$ knowledge conversion process. Compared with the general construction process, the DIK system has the following two advantages: (1) first, the DIK system incorporated the importance assessment of the nodes in the knowledge graph, i.e., redundant data can be removed and important information can be prevented from being lost; (2) second, the medical resource elements are divided into a hierarchy, which can increase the management efficiency in the face of large amounts of data.

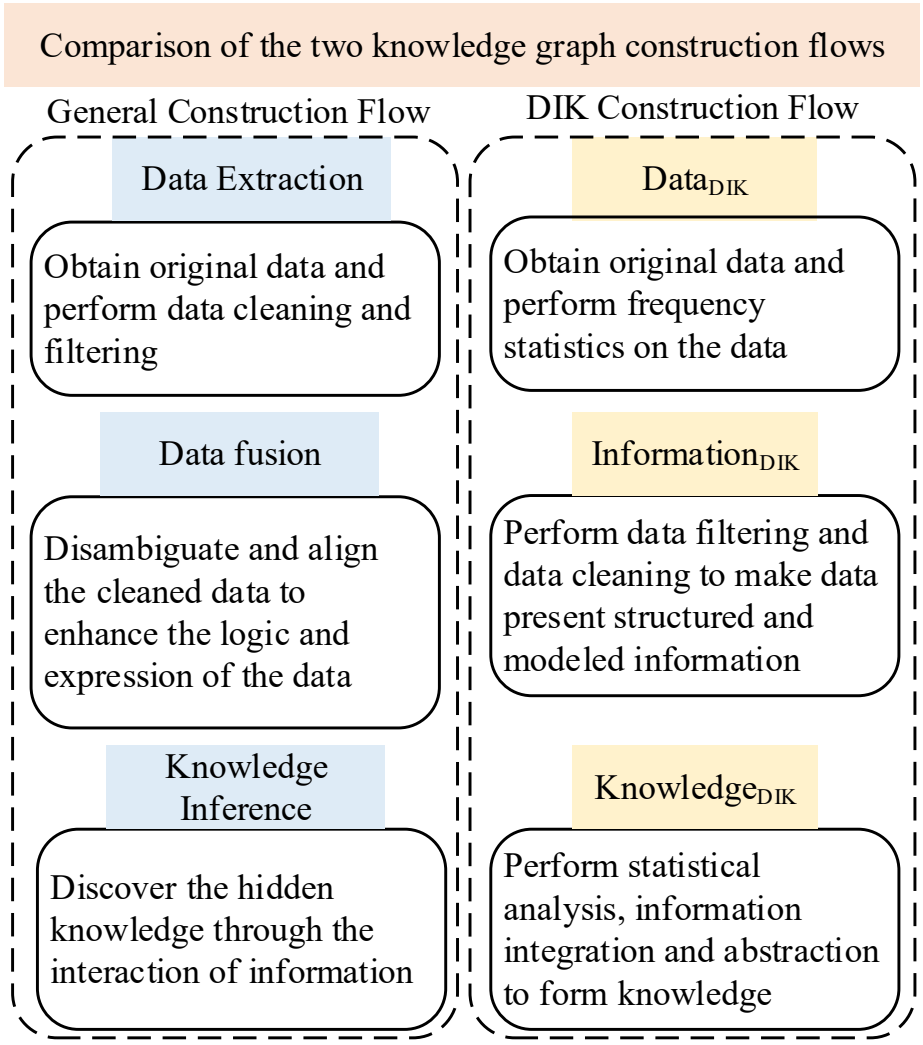

Figure 5. Comparison of two knowledge graph construction processes.

4.3.1. Medical Knowledge Extraction Method

Medical knowledge extraction methods are mainly divided into two ways: manual extraction and automatic extraction. Manual extraction is a sentence of certain rules to collect and organize relevant medical information and extract knowledge, which currently includes ICD-11, clinical medical knowledge base, SNOMED-CT, and so on. Although its accuracy is high, the process of its construction is too time- and energy-consuming. Automatic extraction, on the other hand, uses techniques such as data mining, machine learning, and artificial neural networks to automatically extract essential elements from medical resources. For example, the integrated medical language system UMLS is constructed by the automatic extraction approach.

Currently, BiLSTM-CRF is the most mainstream deep-learning model for entity extraction in the medical field. Jagannatha et al. [13] experimentally compared BiLSTM-CRF with other and its learning models for entity extraction of electronic medical records, and the experimental results showed that the BiLSTM-CRF model was effective in improving the accuracy of the results. Therefore, in the medical knowledge extraction phase of our work, we used the deep learning model BiLSTM-CRF model and combined it with a semi-supervised learning method (Bootstrapping) to complete the data extraction task for Chinese electronic medical records, and the extraction results included four types of medical entities and nine types of entity relationships, and on this basis, we constructed a medical knowledge graph based on electronic medical records. In addition, to further improve the accuracy of medical entity recognition, the Attention Mechanism was introduced,

and a new model BiLSTM-Attended-CRF model was constructed to assign an attention weight to each word so that the model can assign higher weights to entity words, and experiments were designed to demonstrate the effectiveness of the method [14].

Web crawlers as a widely used and automatic way to collect data can obtain the needed data by fetching the specified web pages and parsing them. Feng et al. [15] designed and implemented a distributed crawler-based Web spatial data acquisition system and tested the effectiveness of the system. Pang [16] designed a Python-based distributed crawler system, which mainly includes three modules: a data storage model, web capture module, and web analysis module, and the three modules cooperate to complete the data collection task. We used the existing Scrapy crawler framework to implement the data collection of online medical communities.

4.3.2. Two-Step Data Fusion Implementation

Medical knowledge fusion is founded on medical data extraction, and how to eliminate uncertainty in knowledge understanding, discover the true value of knowledge, and expand the correct knowledge update to the knowledge base is the focus of attention in medical knowledge fusion research [17].

The key techniques of medical knowledge fusion contain an entity alignment technique, entity linking technique, and relationship deduction technique. Entity alignment techniques are used to eliminate the heterogeneity of ontologies and data sources, entity linking is the basis of medical knowledge fusion, and inconsistencies in knowledge are eliminated by operations such as entity disambiguation; the relational inference is used to discover implicit knowledge, thus expanding and complementing the medical knowledge base.

The specific implementation of the two-step data fusion strategy is as follows (as shown in Figure 6):

(a) The first step is the data processing operation. The processing objects were for entity-type data that were obtained from different data sources using different methods. The data were processed using entity alignment methods to eliminate redundancy, remove errors, and perform alignment operations on entity terms with the same meaning and different names extracted from the data sources. When there was an entity alignment conflict, the entity in the medical data source was used as the standard, avoiding the incorrect operation of fusing similar disease entities into the same entity;

(b) The second step is the fusion operation of medical data characteristics. In medical data, the role of disease entities was more central than that of various entities such as symptoms, treatments, drugs, departments, etc. Therefore, this paper utilized the characteristics of the "disease hub" to fuse medical entities from different data sources according to the developed medical relationship type, oriented by the names of disease entities. For example, the disease entity "tuberculosis" in the electronic medical record is firstly combined with the "A15.001" data in the medical dictionary through the "disease code" relationship type to form triplet data. Then, the relationship types of "symptoms", "department", "examination" and "treatment" are combined with the corresponding data in the electronic medical record. Finally, the relationship types of "food" and "medicine" are combined with the corresponding data in the medical community.

In addition, the entity alignment results are shown in Table 4, which count the operation results of the two methods for entity fusion. The difference in the number of disease entities between the method in this paper and the conventional method could directly indicate that the method used in this paper has improved the fusion of disease entities. In addition, the difference in the total number of medical entities between the two methods was equal to the difference in disease entities, which indicated that the method in this paper could effectively avoid the wrong operation of fusing different disease entities without losing the number of medical entities other than diseases.

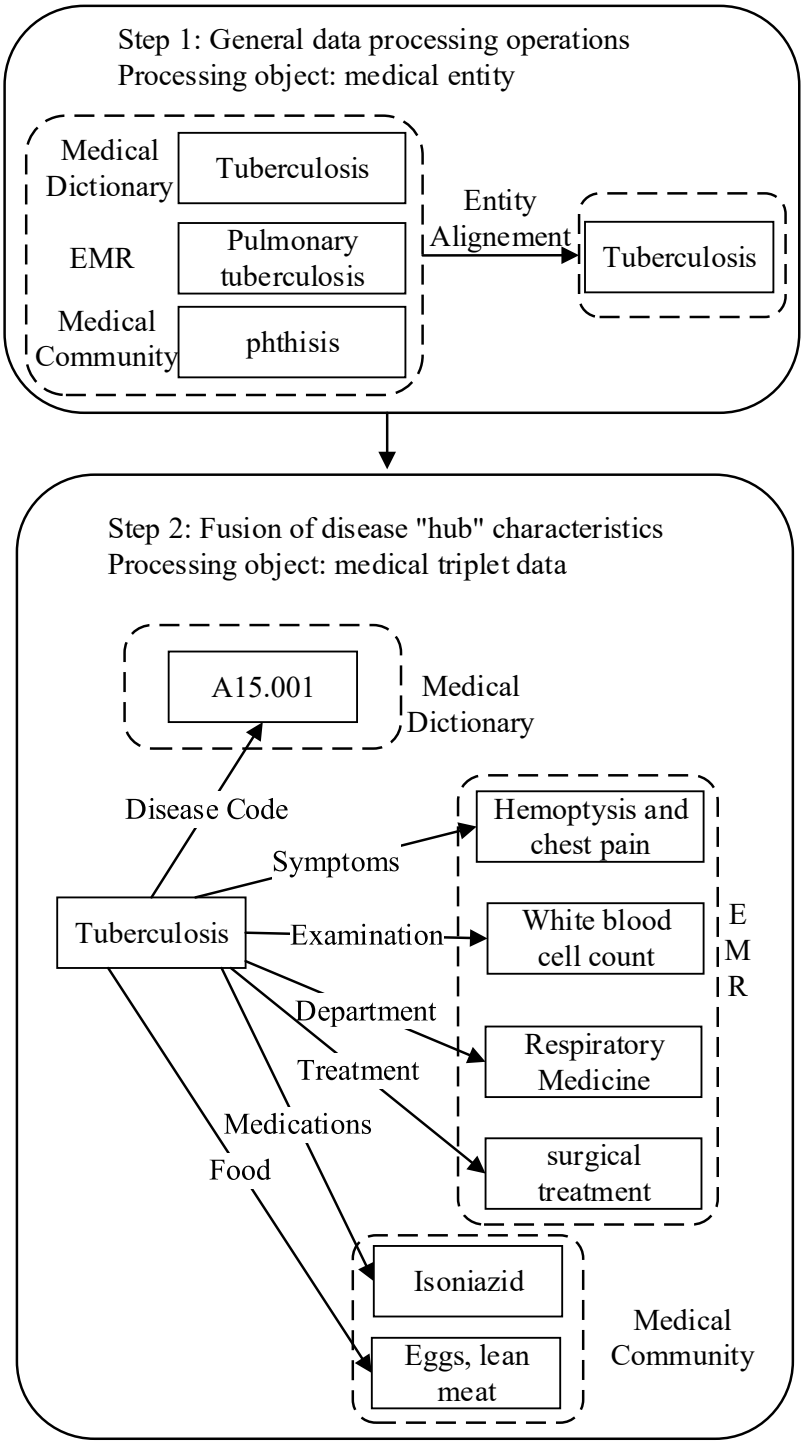

Figure 6. Two-Step Data Fusion Strategy.

Table 4. Statistics of the results of the two entity alignment methods.

Method	Number of Disease Entities	Number of Medical Entities
Solid Alignment (General method)	1275	39,359
Entity Alignment (This paper's method)	1912	41,271

4.3.3. Construction of the DIK Medical Knowledge Graph

The process of building a knowledge graph started from the raw data and used a series of automatic or semi-automatic technical methods to extract knowledge elements (facts) from the raw data and store them in the knowledge base, which was an iterative updating process. Knowledge graph construction methods include top–down and bottom–up. Top–down construction is performed with the help of structured data sources such as encyclopedic websites, from which ontologies and schema information were extracted and added to the knowledge base. Bottom-up is built by using some techniques to extract resource patterns from publicly collected data and then added to the knowledge base after manual review. The multi-data source DIK medical knowledge graph constructed in this paper adopted the bottom–up knowledge graph construction method. Considering the generality of medical knowledge graphs, medical data was stored according to triads (entity1 → entity relationship → entity2), etc., where entity relationship is a directed edge. We imported the collated triad data into the Neo4j database to realize the construction of the DIK medical knowledge graph in the form of directed graph storage.

Figure 7a shows that Neo4j database automatically assigns different colors according to different entities (nodes) and different edges (relationships) within the knowledge graph. There were only three colors in Figure 7b, where yellow nodes indicated disease names, red nodes indicated data from Chinese electronic medical records, and green nodes were data from online medical communities.

(**a**)

Figure 7. *Cont.*

(**b**)

Figure 7. The Medical Knowledge Graph. (**a**). Visualization of DIK Medical Knowledge Graph from Multi-Data Source; (**b**). Medical Data Fusion from Different Data Sources.

5. The Application and Experiment

5.1. The Application of Question and Answer

Knowledge graphs provide a more effective way to express, organize, manage, and utilize massive, heterogeneous, and dynamic medical big data in medical information systems, making the systems more intelligent and closer to human cognitive thinking. Currently, medical knowledge graph technology is mainly used in clinical decision support systems [18], medical intelligent semantic search engines [19], medical question and answer systems [20], chronic disease management systems, medical guidance systems, and adverse drug reactions [21]. Medical question and answer systems were an advanced form of medical information retrieval that could provide users with answers in an accurate and brief natural language form. Abacha et al. [22] proposed a medical question-and-answer system based on natural language processing, which combined medical domain knowledge, natural language processing-related techniques, and semantic relations to construct a knowledge graph for the automated question-and-answer of medical problems. Ruan et al. and Mu [23] integrated TCM and EMR knowledge graphs such as disease database, symptom database, and herbal medicine database, and explored the automated construction method and standardized process of TCM knowledge graphs using text extraction, relational data conversion, and data fusion to realize the intelligent application of TCM knowledge graphs.

In this paper, we used the Flask framework and Python language to construct an intelligent medical Q&A system based on the DIK medical knowledge graph, whose implementation flow is shown in Figure 8, which realized the conversion of "natural language-to-Cypher-to-natural language" (Cypher is the query language of Neo4j). The first "natural-language-to-Cypher" is the conversion of the user-entered question to the system's internal search statement. The second "Cypher-to-Natural Language" is the conversion of the system's internal search results to the user's understandable natural language form of the response. In this paper, the conversion of natural language queries and responses was used to improve the serviceability and friendliness of the intelligent medical Q&A system.

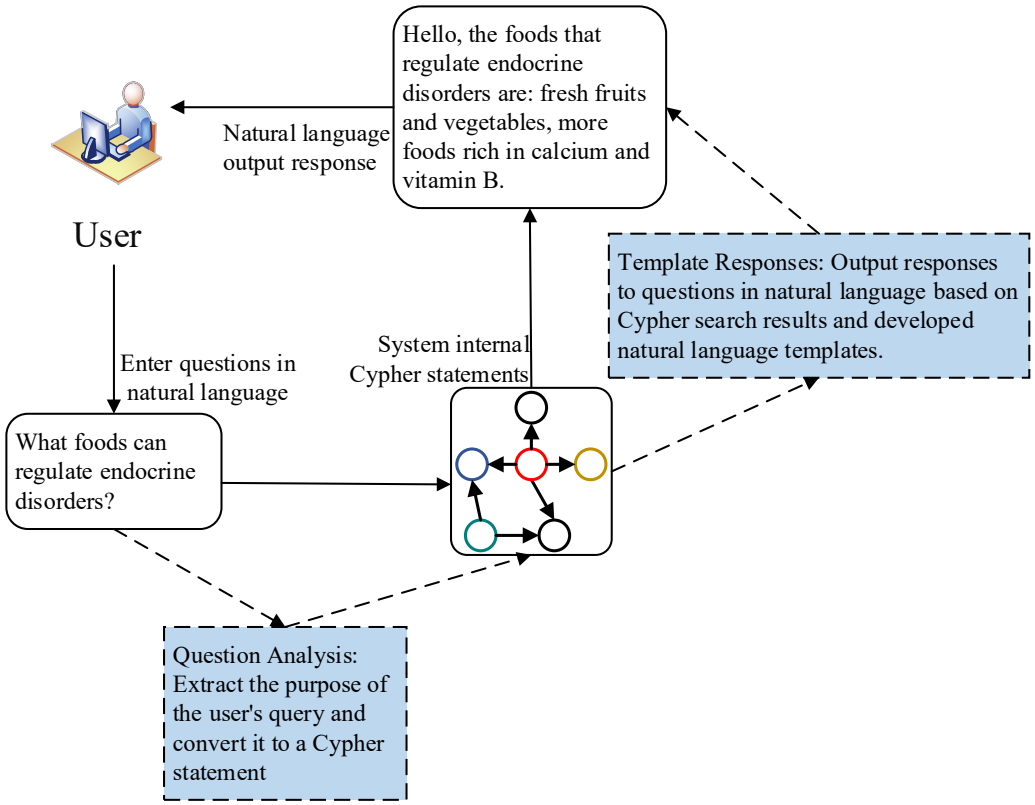

Figure 8. The implementation process of a medical intelligent question-answering system.

5.2. Experiment

To make the advantages of the DIK three-layer knowledge map more intuitive, this paper designed a set of simple comparison experiments by comparing responses to the same medical questions, i.e., the general structure of the medical Knowledge Graph (KG) and DIK three-layer medical Knowledge Graph (KG_{DIK}) were chosen as two knowledge bases under the same condition constraints of data sources. Then, the two knowledge bases were implemented with the intelligent question-and-answer system through the same question base, and the effectivenesses of the two different structured knowledge graph structures were compared by the number of responses and accuracy of the question-and-answer system responses, and the experimental details are shown in Tables 5 and 6. The number of responses and their accuracy rates in Table 6 can illustrate the effectiveness of the second-step data fusion strategy.

Table 5. Q&A dataset statistics.

Type	Quantity
Question	54,000
Answer	101,743

Table 6. The experiment results of intelligent QA by KG and KG_{DIK}.

Knowledge Graph Type	Number of Replies (Pieces)	P (%)	R (%)	F1 (%)
KG	26,894	49.80	47.90	49.13
KG_{DIK}	36,352	85.84	84.72	85.27
KG (two-step)	42,744	58.47	54.92	56.60
KG_{DIK} (two-step)	76,082	85.70	86.05	85.87

The dataset used in this experiment is from a standard dataset provided by a third party [24], which collected professional questions and answers, spoken question and answer types such as online medical communities, doctors' consultations, online science, etc. Two sets of comparative experiments are designed in this section:

(1) The comparison of methods for the construction of knowledge graphs (KG, KG_{DIK});
(2) The comparison of the data fusion strategies with and without (KG, KG (two-step) and KG_{DIK}, KG_{DIK} (two-step)).

As for the comparison of the first set with or without DIK architecture, it can be seen from the number of replies that the number of replies of KG, in general, is significantly lower than that of KG_{DIK}, which indicates that the DIK architecture knowledge graph framework can improve the correlation of data to some extent. Second, in terms of the performance evaluation of the question-and-answer model, it can be seen that the F1 value of KG in general is significantly lower than that of KG_{DIK}. From the evaluation results, it can be concluded that the DIK architecture can not only improve the correlation between data but also improve the precision of data matching and the performance of the method.

About the second set of experimental comparisons with and without data fusion strategy, to clarify the influence of DIK architecture and data fusion strategy on the experiments, this set of comparisons was divided into two sub-groups, i.e., with and without DIK architecture. The first is the impact of the with and without data fusion strategy on the general knowledge graph; the experimental results show that the number of responses has improved and is higher than the experimental results of the first group, indicating the effectiveness of the data fusion strategy. The secondary is the effect of the knowledge graph with and without the data fusion strategy over the KG_{DIK}. Through the results, it can be revealed that the data fusion strategy can effectively increase the relevance and consistency of the data and reduce data loss during data processing.

The second step of the two-step data strategy to merge the fragmented triplet data into a "treatment plan" with the characteristics of a "disease hub" can effectively enhance the overall view of the medical knowledge graph in the question-and-answer system, which in turn can improve the effectiveness of the medical knowledge graph.

6. Discussion

The experiment in Section 5.2 objectively illustrates that the research method of this article, under certain conditions, can achieve accuracy and quickly find the user-demanded data for research purposes. However, the complicated constraints cannot make it well applied in the field.

After this paper, it is hoped to inspire more thinking, for example, is there only one approach to the strategy for knowledge discovery? How effective are other knowledge discovery strategies compared with the DIK knowledge discovery strategy of this paper? For another example, the constraints proposed in the proposed method should be reduced in the subsequent work in order to facilitate implementation. It is necessary to study how

to quickly realize effective data fusion, quickly build knowledge graphs, and quickly use them [25].

Besides, the biggest limitations and innovations of this research are in the proposed two-step data fusion strategy, because it is specifically designed according to the characteristics of medical data out of a data fusion strategy and is not widely used in other data sets. Although there is a certain degree of improvement, compared with the amount of work, it may make people forget its advantages [26–28]. However, it may be valuable to consider from the features of the incorporation data fusion strategy.

Because the current data fusion strategy is for the common characteristics of the data, the role of each field is similar, and cannot effectively make more effective changes for the deep data characteristics [29].

7. Conclusions

In the medical field, with the gradual advancement of the medical informatization level, many medical resources have been accumulated, and the construction of the DIK medical knowledge graph with multiple data sources provides a method to extract knowledge from many medical resources, which has a broad application prospect. In this paper, firstly, according to the DIKW system and knowledge graph construction process, a set of DIK three-layer knowledge graph construction methods conforming to medical knowledge induction was proposed. Secondly, a data source selection strategy was proposed for the case of multiple data sources; again, a two-step data fusion strategy was developed for the characteristics of medical data to improve the effective fusion between different data sources. Finally, based on DIK medical knowledge graph, an intelligent medical question-and-answer system wasis built by combining the Flask framework and Python to realize the "Natural Language Query → DIK Medical Knowledge Graph → Natural Language Answer" process. To evaluate the effectiveness of the DIK medical knowledge graph, a set of simple comparative experiments was designed to verify it. The research approach proposed in this paper, taking into account the characteristics of the method itself and the data, had the purpose of designing a new data strategy and methodological model from a practical application perspective.

This paper mainly discusses the application of DIK architecture in the medical field. However, the DIK architecture itself is a general knowledge framework that can be used for application implementation in various industries. Therefore, the method proposed in this paper still has some defects, shortcomings, and limitations, as shown below:

(a) Inaccuracy of entity correspondence. The main challenge in the medical knowledge fusion phase is to achieve accurate entity linkage. The possible causes are that the diversity of medical knowledge sources leads to serious multi-source referencing problems of medical entities in different data sources [30–32]. Therefore, how to link the extracted entities accurately and correctly to the medical knowledge base in a context-constrained manner is a common concern in the current academic community;

(b) Storage method for the knowledge graph. The ternary representation is widely used and accepted. However, it suffers from problems such as low computational efficiency as the data volume grows. Therefore, representing the semantic information in medical entities as dense low-dimensional real-valued vector methods will be the next research direction [33,34];

(c) The scale of data is still not large enough, although the types of data in this study are diverse. For the DIK architecture to collect richer and more accurate information from the data, the data scale needs to be further expanded in future work [35,36];

(d) The imperfection of the architecture of DIK. Originally, the DIKW was a complete architecture, but due to the planning of the research work, the three-layer architecture of the DIK was adopted in this paper. In order to use the complete DIKW architecture, it is necessary to consider how to proceed from the knowledge level to a higher level of abstracted knowledge [37].

Author Contributions: Conceptualization, M.L.; methodology, M.L.; validation, Z.N.; investigation, Z.N. and L.T.; writing—original draft, M.L.; writing—review and editing, M.L., Z.N., L.T., Y.H., J.S. and Y.W.; funding acquisition, L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (No. 2022YFB2901304), and was funded by the Innovation Scientists and Technicians Troop Construction Projects of Henan Province (No. 224000510002), and was funded by Program of Songshan Laboratory (Included in the management of Major Science and Technology Program of Henan Province) (No. 221100210900-03), and (No. JCKY2018210B022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hou, M.; Wei, R.; Lu, L.; Lan, X.; Cai, H. Research Review of Knowledge Graph and Its Application in Medical Domain. *J. Comput. Res. Dev.* **2018**, *55*, 2587–2599.
2. Huang, M.; Li, M.; Han, H. Research on entity recognition and knowledge graph construction based on electronic medical records. *Comput. Appl. Res.* **2019**, *36*, 3735–3739.
3. Yuan, K.-Q.; Deng, Y.; Chen, D.; Zhang, B.; Lei, K. Construction techniques and research development of medical knowledge graph. *Appl. Res. Comput.* **2018**, *35*, 1929–1936.
4. Liu, C. Reseach of the Medical Knowledge Based on Knowledge Graph. Master's Thesis, Zhejiang Sci-Tech Univeristy, Hangzhou, China, 2017.
5. Zhou, M. *The Research and Development of Question Answering System Based on Knowledge Graphs*; Beijing University of Posts and Telecommunications: Beijing, China, 2017.
6. Wang, H.; Zhang, J.; Cheng, X. Construction of Chinese Open Link Medical Data. *China Digit. Med.* **2013**, *8*, 5–8+15.
7. Han, P.; Ma, J.; Zhang, J.M.; Liu, Y.Z. The framework Construction of Medical Knowledge Graph Based on Multi-data source Fusion. *J. Mod. Inf.* **2019**, *39*, 81–90.
8. Shao, L.X.; Duan, Y.C.; Sun, X.B.; Gao, H.; Zhu, D.; Miao, W. Answering who/when, what, how, why through constructing data graph, information, knowledge graph and wisdom graph. In Proceedings of the 29th International Conference on Software Engineering and Knowledge Engineering, Pittsburgh, PA, USA, 5–7 July 2017; KSI Research Inc.: Pittsburgh, PA, USA, 2017; pp. 1–6.
9. Wu, Y.; Yin, A.; Lin, K.; Yu, X.; Lai, G. Research on Knowledge Graph Construction Method Based on Multi-Data Source. *J. Fuzhou Univ. (Nat. Sci. Ed.)* **2017**, *45*, 329–335.
10. Hu, F.H. *Chinese Knowledge Graph Construction Method Based on Multiple Data Sources*; East China University of Science and Technology: Shanghai, China, 2015.
11. Cowie, J.; Lehnert, W. *Information Extraction*; Springer: Berlin/Heidelberg, Germany, 2004.
12. Shao, L.; Duan, Y.; Zhou, C.; Gao, H.; Chen, S. Design of Recommendation Services Based on Data, Information and Knowledge Graph Architecture. *J. Front. Comput. Sci. Technol.* **2019**, *13*, 214–225.
13. Jagannatha, A.N.; Yu, H. Structured prediction models for RNN based sequence labeling in clinical text. In Proceedings of the 2016 Conference on Empirical labeling Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; ACL: New York, NY, USA, 2016; pp. 856–865.
14. Li, M.; Zhang, Y.; Huang, M.; Chen, J.; Feng, W. Named Entity Recognition in Chinese Electronic Medical Record Using Attention Mechanism. In Proceedings of the 12th IEEE International Conference on Cyber Physical and Social Computing, Atlanta, GA, USA, 14–17 July 2019.
15. Feng, L.; Huang, L.; Zeng, L.; Zhu, Q. Research on Web Spatial Data Acquisition Based on Distributed Web Crawler. *J. Guizhou Univ. (Nat. Sci.)* **2019**, *36*, 33–36.
16. Pang, F. Design and Implementation of Distributed Web Crawler System Based on Python. *Electron. Technol. Softw. Eng.* **2018**, *23*, 6.
17. Dong, X.L.; Cabrilovich, E.; Heitz, G.; Horn, W.; Murphy, K.; Sun, S.; Zhang, W. From data fusion to knowledge fusion. *Proceeding VLDB Endow.* **2014**, *7*, 881–892. [CrossRef]
18. Carcia-Cresp Rodriguez, A.; Mencke, M.; Gómez-Berbís, J.M.; Colomo-Palacios, R. ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. *Expert Syst. Appl.* **2010**, *37*, 2621–2628.
19. Huang, C.C.; Liu, Z. Exploring query expansion for entity searches in PubMed. In Proceedings of the 7th International Workshop on Health Text Mining and Information Analysis, Austin, TX, USA, 5 November 2016; pp. 106–112.
20. Terol, R.M.; Martinez-Barco, P.; Palomar, M. A knowledge based method for the medical question answering problem. *Comput. Biol. Med.* **2007**, *37*, 1511–1521. [CrossRef]

21. Mou, D.; Ju, Y.; Dai, W.; Huang, L. Knowledge Discovery Strategy and Model of Virtual Health Community Text Data. *Libr. Inf. Serv.* **2018**, *62*, 125–130.
22. Abacha, A.B.; Zweigenbaum, P. MEANS: A medical question-answering system combing NLP techniques and semantic Web technologies. *Inf. Process. Manag.* **2015**, *51*, 570–594. [CrossRef]
23. Ruan, T.; Sun, C.; Wang, H.; Fang, Z.; Yin, Y. Construction of traditional Chinese medicine knowledge graph and its application. *J. Med. Inform.* **2016**, *37*, 8–13.
24. Mu, Y.Z. *Research on Chinese Electronic Medical Record Entities Recognition and Entity Relation Extraction Based on Semi-Supervised Learning*; Hainan University: Haikou, China, 2018.
25. Ji, G.; Liu, K.; He, S.; Zhao, J. Distant Supervision for Relation Extraction with Sentence-Level Attention and Entity Descriptions. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 3060–3066.
26. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489.
27. Yang, P.; Yang, Z.; Luo, L.; Lin, H.; Wang, J. An Attention-Based Approach for Chemical Compound and Drug Named Entity Recognition. *J. Comput. Res. Dev.* **2018**, *55*, 1548–1556.
28. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, 7–12 August 2016; Volume 2, pp. 207–212.
29. Fang, T. *Medical Knowledge Map Construction Based on Chinses Language Processing and Deep Learning*; Henan Normal University: Xinxiang, China, 2018.
30. Zhong, L. Research on the Construction Method of Chemical Knowledge Map for Baidu Encyclopedia. *Softw. Guide* **2017**, *16*, 168–170.
31. Azeem, M.; Jamil, M.K.; Shang, Y. Notes on the Localization of Generalized Hexagonal Cellular Networks. *Mathematics* **2023**, *11*, 844. [CrossRef]
32. Nadeem, M.F.; Azeem, M. The fault-tolerant beacon set of hexagonal Möbius ladder network. *Math. Meth. Appl. Sci.* **2023**, 1–15. [CrossRef]
33. Zhang, X.; Kanwal, M.; Azeem, M.; Jamil, M.; Mukhtar, M. Finite vertex-based resolvability of supramolecular chain in dialkyltin. *Main Group Met. Chem.* **2022**, *45*, 255–264. [CrossRef]
34. Raza, H.; Sharma, S.K.; Azeem, M. On Domatic Number of Some Rotationally Symmetric Graphs. *J. Math.* **2023**, *2023*, 3816772. [CrossRef]
35. Azeem, M.; Imran, M.; Nadeem, M.F. Sharp bounds on partition dimension of hexagonal Möbius ladder. *J. King Saud Univ.-Sci.* **2022**, *34*, 101779. [CrossRef]
36. Shabbir, A.; Azeem, M. On the Partition Dimension of Tri-Hexagonal α-Boron Nanotube. *IEEE Access* **2021**, *9*, 55644–55653. [CrossRef]
37. Azeem, M.; Nadeem, M.F. Metric-based resolvability of polycyclic aromatic hydrocarbons. *Eur. Phys. J. Plus* **2021**, *136*, 395. [CrossRef]