



DEKGB: An Extensible Framework for Health Knowledge Graph

Ming Sheng¹, Yuyao Shao², Yong Zhang^{1(✉)}, Chao Li¹,
Chunxiao Xing¹, Han Zhang³, Jingwen Wang², and Fei Gao⁴

¹ RIIT&BNRCIST&DCST, Tsinghua University, Beijing 100084, China
{shengming, zhangyong05, lichao, xingcx}@tsinghua.edu.cn

² Beijing Foreign Studies University, Beijing 100089, China
{shaoyuyao, wjw}@bfsu.edu.cn

³ Beijing University of Posts and Telecommunications, Beijing 100876, China
zhanghan3281@bupt.edu.cn

⁴ Henan Justice Police Vocational College, Zhengzhou 450046, China
68521617@qq.com

Abstract. With the progress of medical informatization and the substantial growth of clinical data, knowledge graph is playing an increasingly important role in medical domain. Medical domain is highly specialized with abundant high-quality ontologies, and has many professional sub-fields such as cardiovascular diseases, diabetes mellitus and so on. It is very difficult to build a health knowledge graph for all of the diseases because of data availability and deep involvement of doctors. In this paper, we propose an efficient and extensible framework, DEKGB, to construct knowledge graphs for specific diseases based on prior medical knowledge and EMRs with doctor-involved. After that, we present the detailed process how DEKGB is applied to extend an existing health knowledge graph to include a new disease. It is confirmed that using this framework, doctors can get highly specialized health knowledge graphs conveniently and efficiently.

Keywords: Health knowledge graph · Extensibility · Conceptual graph · Instance graph

1 Introduction

Knowledge graphs have been in the focus of research since 2012 when the term was firstly proposed by Google, as an enhancement of their search engine with semantics [1]. A popular definition of knowledge graph is a graph-based data structure composed of nodes (entities) and labeled edges (relationships between entities) [2]. In general, we define knowledge graphs with only concept nodes as Concept Knowledge Graph (CKG), knowledge graphs with both instance nodes and event nodes as Instance Knowledge Graph (IKG), and knowledge graphs covering CKG and IKG as Factual Knowledge Graph (FKG).

Current health knowledge graphs usually cover wide areas of medical knowledge: all proteins (UniProt), as many drugs as possible (Drugbank), as many drug-drug

interactions as are known (Sider), and massively integrated knowledge graphs such as Bio2RDF and LinkedLifeData [3]. However, the knowledge graphs covering wide areas may contain inaccurate and irrelevant knowledge so as to be difficult in actual use. For user-friendly and accurate purpose, health knowledge graphs should be tailored for doctors in specific diseases. Meanwhile another problem may occur. Building a new knowledge graph for a new disease from scratch in medical domain is time consuming because there are many repetitive procedures such as data alignment. So a construction framework for knowledge graph in a specific disease is necessary. Currently, the construction framework for general knowledge graphs mainly includes knowledge representation, knowledge graph building tools such as extraction tools, and knowledge storage and application. However, general strategies cannot be directly applied to domain-specific knowledge graphs, let alone to more sophisticated disease-specific knowledge graphs. The reason is that specific entities and relations from specific data sources need to be extracted, and specialized semantic networks for different diseases need to be constructed. In medical domain, there are many complex concepts and relations, diverse diseases, which require large amount of prior knowledge from doctors to clarify them. What's more, the doctors' actual demands may be various. So the help from doctors also plays an important role. Hence, the most important aspects of building a disease-specific knowledge graph lie in three parts: the disease-specific data sources, the building tools to extract specific entities and relations, and the help from the doctors.

The problem we are going to address in this paper is how to construct a knowledge graph for a specific disease and extend it to other diseases, based on prior medical knowledge, EMRs and doctors.

To solve this problem, we propose a knowledge graph building framework DEKGB that can be used to create a knowledge graph with many diseases. We use EMRs from hospitals and implement a toolset to help the doctors put forward professional knowledge, along with existing medical knowledge.

This paper is organized as follows. In Sect. 2 we introduce the related work. In Sect. 3 we present the framework and data flow of DEKGB. In Sects. 4 and 5 we show the construction of CKG and IKG for cardiovascular diseases in detail. In Sect. 6, we show the extension process to include a new disease in an existing health knowledge graph by using DEKGB. In the end, we summarize the paper and propose future work in Sect. 7.

2 Related Work

We investigate several knowledge graph building frameworks in health domain, such as cTAKES, pMineR, I-KAT and RDR. According to the three most important aspects to build disease-specific knowledge graphs in the introduction, the related works will be compared from 3 aspects, data sources, building tools adopted and the help from experts (Table 1).

In general, pMineR supports Processing Mining for clinical data from both administrative and clinical aspect. It provides automatically identification services for process discovery [4] and is currently exploited in Hospitals for supporting domain

Table 1. Comparisons of building frameworks.

Name	Data Sources	Building Tools						The help from experts	Extensibility
		Entity extraction	Relation extraction	Event Extraction	Entity alignment	ER/OWL mapping	Human		
cTAKES	UMLS	√	√		√			√	
pMineR	EMRs,...			√					
I-KAT	SNOMED CT,...					√	√	√	√
RDR							√		
DEKGB	UMLS, EMRs,...	√	√	√	√	√	√	√	√

experts in the analysis of the extracted knowledge models. cTAKES [5] is an open-source Natural Language Processing (NLP) system that extracts clinical information from EMRs. I-KAT provides a user-friendly environment to create Arden Syntax MLM (Medical Logic Module) as shareable knowledge rules for intelligent decision-making by CDSS [6].

2.1 Data Sources of Frameworks

The building frameworks listed above have different data sources. For example, cTAKES and I-KAT collect data from medical databases like UMLS, SNOMED CT and so on, while pMineR gathers data from open source EMRs. In general, the data sources of building frameworks in medical domain are mainly from public resources, including medical standards and clinical records.

2.2 Building Tools of Frameworks

Different frameworks adopt different measures to process massive medical knowledge or clinical data, on the purpose of meeting multiple requirements. The building tools listed above show the differences of building health knowledge graphs.

pMineR can encode clinical events by extracting processes under the form of directed graphs, which can calculate the real model of the processes. It also provides graphical comparison tool between different processes, allows doctors to model the adherence to a given clinical guidelines and to estimate performance together with the workload of the available resources in health care. cTAKES offers Natural Language Processing tools like annotation system to extract entities and relations from EMRs. I-KAT creates a knowledge base from MLMs using Arden Syntax to achieve shareability, uses standard data models and terminologies to enhance interoperability, and reduces complexity by abstraction at the application layer to provide physician friendliness.

In DEKGB, multiple tools are adopted to construct the disease-specific knowledge graphs, including entity, relation and event extraction tool, normalization tool, ER-OWL mapping tool and doctor-involved tools. The tools we adopt support better knowledge extraction and knowledge graph construction.

2.3 The Help from Doctors

Experts offer their prior knowledge and demands to construct health knowledge graphs for different usages and applications. cTAKES offers the creation of a personalized dictionary from UMLS according to experts' demands, to process clinical notes and identify types of clinical named entities - drugs, diseases/disorders, signs/symptoms, anatomical sites and procedures; while I-KAT provides a user-friendly platform for doctors to create knowledge bases based on their prior knowledge and use standard syntax to share the knowledge.

DEKGB supports doctors to input personalized prior knowledge. At the same time, public medical standards will also be involved, to construct a disease-specific CKG, meeting experts' demands and covering comprehensive medical knowledge simultaneously. DEKGB also allows experts to propose extraction rules on professional medical records to form instance knowledge graphs that can exploit in-depth medical knowledge and their associations. What's more, for both CKG and IKG, DEKGB offers tools for experts setting different mapping rules to convert structured data to triples in knowledge graphs.

3 Framework of DEKGB

In this section, we will present the framework of DEKGB. The framework includes framework architecture and work flow of DEKGB in detail.

3.1 Architecture of DEKGB

DEKGB can build a disease-specific knowledge graph or can be applied to an existing disease-specific knowledge graph to extend the single disease to other diseases, such as extending from cardiovascular diseases to diabetes mellitus. It is suitable for expanding to all diseases incrementally. As shown in Fig. 1, based on the existing medical thesaurus, EMRs from hospitals in a specific disease and doctors in the study of that disease, the construction of a disease-specific knowledge graph could be implemented.

The building process of knowledge graphs using DEKGB can be divided into two modules:

1. CKG Building Module: construction of concept knowledge graphs from doctors' prior knowledge and medical standards like UMLS.
2. IKG Building Module: extraction of entities and relations from EMRs and fusion of concept nodes and instance nodes.

Specifically, DEKGB introduces doctor-involved tools into building modules, including doctor input tool, rule base tool and doctor annotation tool.

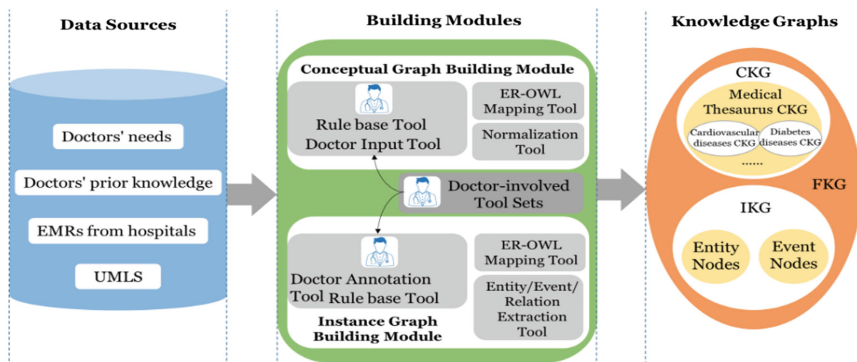


Fig. 1. The framework of DEKGB.

3.2 Work Flow

As shown in Fig. 2, three main problems that DEKGB needs to solve are data sources, building modules and the help from doctors.

The data sources of DEKGB are (1) EMRs from disease-specific Hospital: clinical data from professional hospitals specialized in a specific disease, (2) doctors in related diseases: leading doctors in the study of the specific disease and (3) medical thesaurus containing different medical standards like UMLS. To involve the doctors in DEKGB, we propose doctor-involved tools: doctor input tool, doctor annotation tool and rule base tool. Doctor input tool and rule base tool are used in the construction of CKG while doctor annotation tool and rule base tool are applied to build IKG. For building modules, the construction of CKG is based on doctors' prior knowledge and medical thesaurus and this module contains normalization tool, ER-OWL tool and doctor-involved tools. Meanwhile the building modules of IKG are divided into structured data conversion module (from ER model to RDF model) and unstructured data conversion module (extraction of entities, relationships and events). And the construction of IKG is based on EMRs from hospitals, extraction tools and doctor-involved tools. Finally, for different possible usages, DEKGB generates CKG and IKG.

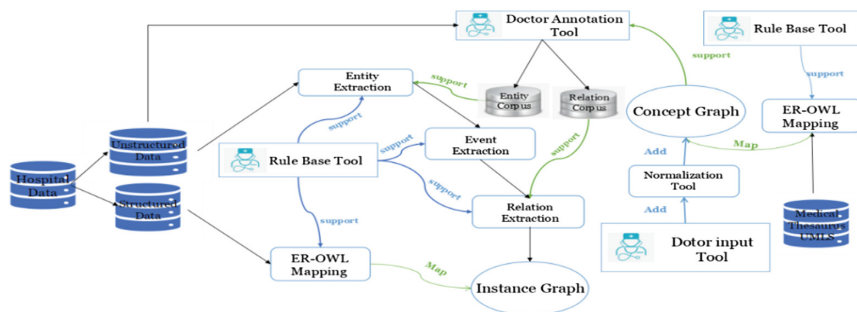


Fig. 2. The data sources and building workflow of DEKGB.

4 Building CKG for Cardiovascular Diseases

Conceptual knowledge graph integrates medical standards from medical thesaurus like UMLS and doctors' prior knowledge from the doctor input tool. Two steps are contained in the conceptual graph building module. The first step is the construction of medical thesaurus CKG and it will be conducted when DEKGB is firstly applied to construct a disease-specific knowledge graph. The second step is the construction of a disease-specific CKG when a new disease is going to be included in existing CKG. Here we take building CKG for cardiovascular diseases as an example (Fig. 3).

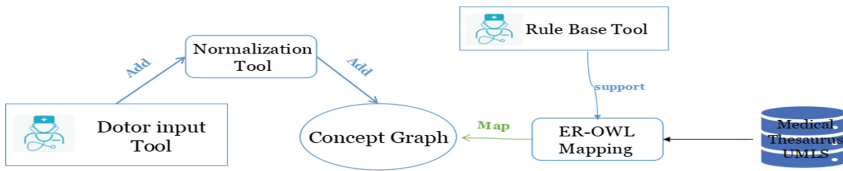


Fig. 3. Conceptual graph building module.

4.1 Medical Thesaurus CKG Construction Procedure

In DEKGB, the data source of Medical thesaurus CKG is UMLS. UMLS is the abbreviation of Unified Medical Language System, a set of files and software that brings together many health-related standards. The medical knowledge in ULMS is stored in ER databases. Hence, medical knowledge in UMLS needs to be mapped to nodes in CKG. For user-friendly purpose, DEKGB provides a mapping rule tool for doctors to decide the mapping format. The generated mapping rules in rule base is applied to convert the data from ER database to RDF graphs [7]. In specific, doctors set rules indicating which column (in ER databases) should convert to which concept node (in knowledge graphs).

At present, the conversion from ER to RDF includes direct mapping and custom mapping. Here we adopt a custom mapping method to better meet doctors' needs. The standards and tools that support this method are as follows: R2RML, Virtuoso, etc. In DEKGB, we use R2RML, a standard transformation language of converting a relational database to RDF and its implement tool D2R is adopted to fulfill the process.

4.2 Cardiovascular Diseases CKG Construction Procedure

The disease-specific CKG is constructed on the basis of doctors' prior knowledge and medical thesaurus CKG. To get doctors' prior knowledge, DEKGB provides the doctor input tool for doctors to define concepts and relations in the new disease field and add

them into CKG. Here we take cardiovascular diseases for example. The steps to construct cardiovascular diseases CKG are as follows.

1. A group of doctors in cardiovascular diseases are invited to offer crucial medical knowledge through doctor input tool. The prior knowledge from doctors contains medical entities, relations and triples, which may be useful for diagnosis or other actual usage.
2. The knowledge from doctors are put into normalization tool. The process of normalization is divided into two kinds, depending on whether the concepts, relations or triples offered by doctors exist in medical thesaurus CKG or not. If the knowledge does not exist, it will be inserted into medical thesaurus CKG according to the encoding system of UMLS. Otherwise, standard medical knowledge encodes by UMLS will be used. Notice that for efficiency, the standard encoding is merely an attribute of the Disease-specific CKG nodes.

Hence, a comprehensive medical thesaurus spanning all kinds of diseases can be constructed incrementally. What's more, despite a universal medical thesaurus CKG, small scale of disease-specific conceptual knowledge graphs will be constructed individually, with which doctors' actual needs can be satisfied. Table 2 shows different normalization formats for different kinds of doctors' prior knowledge after knowledge normalization in cardiovascular diseases.

Table 2. Knowledge normalization in cardiovascular diseases.

Prior Knowledge	In medical thesaurus CKG or not	CUI (concept unique identifier)	LAT (language type)	...	AUI (atom unique identifier, may be various)	SAB (Source Abbreviation, may be various)	STR
unstable angina	√	C0002965	ENG	...	A22810054	SNOMEDCT_US	Unstable angina
					A18571195	CHV	
					
acute cerebral infarction	×	C010523	ENG	...	A999010523	Doc-Def	acute cerebral infarction
原发性高血压3级 (极高危)	×	C020379	CHN	...	A999020379	Doc-Def	原发性高血压3级 (极高危)

5 Building IKG for Cardiovascular Diseases

To build the IKG in an assigned disease field effectively, structured clinical data and unstructured clinical data should adopt different conversion procedures. The procedures include: (1) structured data to IKG and (2) unstructured data to IKG. Here we take the building of IKG for cardiovascular diseases for example (Fig. 4).

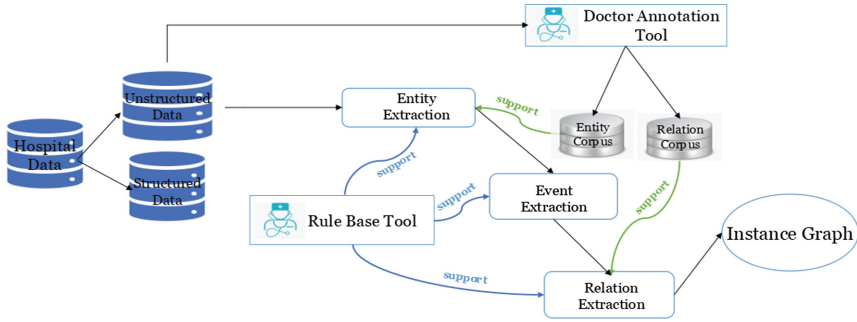


Fig. 4. Instance graph building module.

5.1 Structured Data Conversion Procedure

The main process of structured data transformation is mapping from structured data in ER databases to RDF/OWL instance knowledge graphs, supported by mapping rules from doctors setting through rule base tool. The detailed technology is introduced in the conversion of UMLS to medical thesaurus CKG in Sect. 4.1.

5.2 Unstructured Data Conversion Procedure

The key tools of unstructured data conversion procedure in DEKGB are as follows: entity extraction [8], event extraction and relation extraction shown in Fig. 5. The purpose of this procedure is to extract entities, events and relations in unstructured data and convert them to IKG. In extraction tasks, machine learning-based methods are widely used, but they may be so noisy as to provide many wrong results. So rule-based data integration and extraction approaches are adopted and it has better interpretability and effective interactive debugging [9]. Hence, for better extraction performance and the ability to process massive EMR data, DEKGB provides rule base tool and doctor annotation tool to involve doctors and adopt machine learning-based methods at the same time. Doctor Annotation Tool provides the annotation interface, with which doctors can annotate the unstructured data based on the concepts provided by the CKG. The annotated results will be stored in the entity corpus and relation corpus, used as training set to support entity and relation extractions with machine learning algorithms. The combination of doctors' annotation and machine learning methods can augment the performance of entity and relation extraction [10].

Entity Extraction Tool. The entity extraction tool in DEKGB is based on the sequence annotation method and the rule-based method. To implement the methods, rule base tool and entity corpus are adopted to support it (Table 3).

Sequence Annotation Method. Based on sequence annotation method, entities and relations can be extracted, and used as training set for machine learning method to extract entities and relations in a larger data set. Comparing the results and efficiency among different models, we use LSTM-CRF and CRF in our method [11, 12]. Here is an example of annotation-based entity extraction in cardiovascular diseases:

Table 3. Annotation-based entity extraction in cardiovascular diseases.

Input:
When the patient got up in a morning 1 week ago, he had **chest pain** with no obvious inducement, located in the **precordial region**, **radiating** to left **shoulder** and left **elbow joint**, accompanied by **sweating** and **palpitations**....

Output:
<chest pain, CUI, C0008031 >
<precordial region, CUI, C0230134>
<radiating, CUI, C0332301>
<shoulder, CUI, C0037004 >
<elbow joint, CUI, C0013770 >
<sweating, CUI, C0038990>
<palpitation, CUI, C0030252>

Pattern-Based Entity Extraction. Pattern-based entity extraction [13] can extract entity and relation at the same time, which will be described in relation extraction.

Relation Extraction. The relation extraction is composed of two parts: pattern-based module and supervised learning method-based module [14].

Pattern-based method consists of two steps: (1) recognition of medical entities and (2) identification of the correct semantic relation between each pair of entities. In DEKGB, patterns are defined by doctors in the rule base. For example, many EMRs in inpatient medical records have patterns like “The patient has symptoms such as ***” and “The patient was admitted because of ***”. Such patterns can be applied to entity extraction and relation extraction. At the meantime, supervised learning method is adopted, which would improve the accuracy when data scale is larger (Table 4).

Here is an example of pattern extraction using DEKGB in cardiovascular diseases:

Table 4. Pattern extraction in cardiovascular diseases.

Input:
The patient had chest tightness after a physical activity 4 years ago, no chest pain, no abdominal pain or diarrhea, no nausea or vomiting, and was diagnosed as "coronary heart disease" in the local hospital.

PATTERN1: The patient/ He / She had + * (may split by ‘,’) + after+ *
Output Entity:
{chest tightness}: C0232292, {physical activity}: C0026606
Output Relation:
Patient<has>chest tightness, chest tightness <after>activity

PATTERN2: no +*+ or + * + , (split by ‘,’)
Output Entity:
{chest pain}: C0008031, {abdominal pain}: C0000737.....
Output Relation:
Patient<no>chest pain, Patient<no>abdominal pain.....

6 Extension to Include a New Disease

In addition to building disease-specific knowledge graphs, DEKGB is also extensible to include new diseases into current knowledge graphs. Except for the input of EMRs and doctors' prior knowledge in another disease, the building tools and the input of medical thesaurus can be reused. Hence, workload of building health knowledge graphs for a specific disease will be reduced and a health knowledge graph covering different kinds of diseases can be constructed incrementally. For example, when DEKGB is going to include diabetes mellitus into the existing cardiovascular diseases knowledge graph, the following steps need to be implemented:

Firstly, DEKGB needs EMRs and doctors specialized in diabetes mellitus. Specifically, doctors need to provide their prior knowledge from three aspects:

- (1) Diabetes mellitus related concepts, relations and RDF triples, and Table 5 presents an example from doctors in C2 hospital;

Table 5. Concepts, relations and triples in diabetes mellitus.

	Doctors' Prior Knowledge
Concepts	Insulin-dependent diabetes mellitus, Hyperglycemia, Mumps virus, Coronary heart disease, Neuropathy, Polyphagia, Polyuria, Polydipsia, Insulin, Vomit.....
Relations	Has_sign_or_symptom, Associated_disease, Cause_of, Diagnosed_by.....
Triples	<Insulin-dependent diabetes mellitus, Cause_of, Hyperglycemia>; <Insulin-dependent diabetes mellitus, Associated_disease, Neuropathy>; <Insulin-dependent diabetes mellitus, Has_sign_or_symptom, Polyphagia>.....

- (2) Mapping rules from ER to RDF based on the structured data in EMRs;
- (3) Entity and relation extraction rules based on the features of unstructured data in EMRs. Table 6 shows an example.

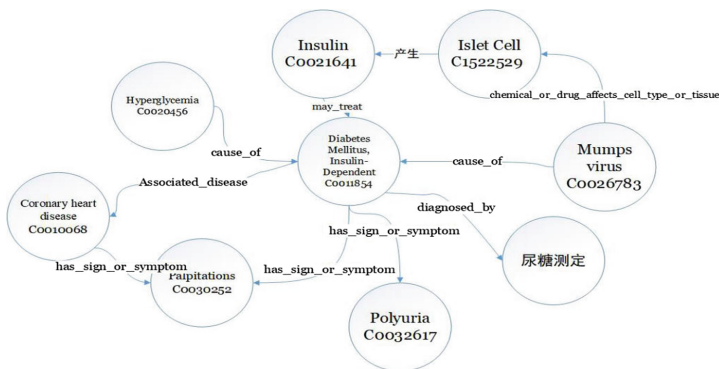
Comparing to cardiovascular diseases, the extraction patterns of diabetes mellitus have their own features, which also proves the necessity of doctors.

Secondly, CKG of diabetes mellitus will be constructed. After getting prior knowledge from doctors through doctor input tool, the concepts, relations and triples will be put into normalization tool for standardization, based on the existing medical thesaurus CKG. Afterwards, the CKG of diabetes mellitus will be implemented. Figure 5 shows a part of CKG of diabetes mellitus based on prior knowledge from doctors after the normalization process.

Finally, IKG of diabetes mellitus will be built. In this process, doctors firstly need to annotate the unstructured data based on the CKG and EMRs in diabetes mellitus. Then machine learning methods are adopted to process massive EMR data and generate entity and relation corpus. Afterwards, on the basis of extraction rules in rule base and automatic extraction methods, the medical knowledge in EMRs of diabetes mellitus will be extracted. Hence, the IKG of diabetes mellitus can be implemented.

Table 6. Pattern extraction in diabetes mellitus.

<p>Input: The patient has found to have hyperglycemia for more than 3 years. The patient was admitted in our department with "diabetes mellitus ". Since onset, the patient has polydipsia, polyphagia, polyuria and 3 kg loss in body weight.</p>
<p>PATTERN1: The patient/ He / She had symptoms such as +*, (split by ‘,’) Output Entity: {headache}: Concept node: C0018681 {dizziness}: Concept node: C0012833/ C0220870/C0012833 /C0149746 {thirst}: Concept node: C0039971 Output Relation: Patient<has>headache Patient<has>dizziness Patient<has>thirst</p>
<p>PATTERN2: Since onset, the patient has ++* (split by ‘,’) Output Entity: {polydipsia}: Concept node: C0085602 {polyphagia}: Concept node: C0020505 {polyuria}: Concept node: C0032617 {body weight}: Concept node: C0005910</p>

**Fig. 5.** CKG of diabetes mellitus.

7 Conclusion and Future Work

In this paper, we propose a disease-specific and extensible knowledge graph building framework named DEKGB. This framework can be used to construct a disease-specific knowledge graph or applied to extend the current health knowledge graph to include a new disease based on medical standards, doctors' prior knowledge and EMRs from a professional hospital. In order to augment the accuracy of knowledge and provide a user-friendlier application, we adopt doctor-involved tools in DEKGB.

In the future, we will gradually involve more medical knowledge from multiple international medical standards and from other channels other than hospitals, to enrich the knowledge graphs built by DEKGB and we will refine the tools we adopt to better fit the actual needs and augment the efficiency and accuracy.

Acknowledgements. This work was supported by NSFC (91646202), National Key R&D Program of China (2018YFB1404400, 2018YFB1402700).

References

1. Ehrlinger, L., Wöß, W.: Towards a definition of knowledge graphs. In: SEMANTiCS (Posters, Demos, SuCCESS) (2016)
2. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs: from multi-relational link prediction to automated knowledge graph construction. CoRR abs/1503.00759 (2015)
3. Huang, Z., Yang, J., van Harmelen, F., Hu, Q.: Constructing disease-centric knowledge graphs: a case study for depression (short version). In: ten Teije, A., Popow, C., Holmes, John H., Sacchi, L. (eds.) AIME 2017. LNCS (LNAI), vol. 10259, pp. 48–52. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59758-4_5
4. Gatta, R., et al.: Generating and comparing knowledge graphs of medical processes using pMineR. In: Corcho, Ó., et al. (eds.) K-CAP, pp. 36:1–36:4. ACM (2017)
5. Savova, G.K., et al.: Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. *J. Am. Med. Inform. Assoc.* **17**(5), 507–513 (2010)
6. Ali, T., et al.: Multi-model-based interactive authoring environment for creating shareable medical knowledge. *Comput. Methods Programs Biomed.* **150**, 41–72 (2017)
7. Unbehauen, J., Martin, M.: SPARQL update queries over R2RML mapped data sources. In: Eibl, M., Gaedke, M. (eds.) GI-Jahrestagung, pp. 1891–1901. GI (2017)
8. Zhang, Z., Tong, Z., Yu, Z., Yali, P.: Attention-based deep residual learning network for entity relation extraction in Chinese EMRs. *BMC Med. Inf. Decis. Making* **19**–S(2), 171–177 (2019)
9. Huang, Z., Wei, X., Kai, Y.: Bidirectional LSTM-CRF models for sequence tagging (2015). <http://arxiv.org/abs/1508.01991>
10. Yimam, S.M., Biemann, C., Majnaric, L., Šabanović, Š., Holzinger, A.: Interactive and iterative annotation for biomedical entity recognition. In: Guo, Y., Friston, K., Aldo, F., Hill, S., Peng, H. (eds.) BIH 2015. LNCS (LNAI), vol. 9250, pp. 347–357. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23344-4_34
11. Fan, J., Li, G.: Human-in-the-loop rule learning for data integration. *IEEE Data Eng. Bull.* **41**(2), 104–115 (2018)
12. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. CoRR abs/1603.01360 (2016)
13. Chen, C., Liu, H., Wang, G., Ding, L., Yu, L.: Entity relationship extraction based on potential relationship pattern. In: Wang, F.L., Gong, Z., Luo, X., Lei, J. (eds.) WISM 2010. LNCS, vol. 6318, pp. 370–377. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16515-3_46
14. Qu, M., Xiang, R., Yu, Z., Jiawei, H.: Weakly-supervised relation extraction by pattern-enhanced embedding learning. In: Champin, P.-A., et al. (eds.) WWW, pp. 1257–1266. ACM (2018)