



How to Define a Multi-modal Knowledge Graph?

Nan Wang¹, Hankiz Yilahun^{2(✉)}, Askar Hamdulla³, and ZhenXuan Qiu¹

¹ Xinjiang University, Urumqi 830017, Xinjiang, China

² School of Information Science and Engineering, Urumqi 830017, Xinjiang, China
hansumuruh@xju.edu.cn

³ Xinjiang Key Laboratory of Multilingual Information Technology, Urumqi 830017,
Xinjiang, China
askar@xju.edu.cn

Abstract. As a form of structured human knowledge, knowledge graphs (KG) have attracted great attention from both the academic and industrial communities since their emergence. It is widely used in the field of artificial intelligence for applications such as information retrieval, data analysis, intelligent question-answering and recommendation systems. In recent years, various types of information on the internet have exploded in growth. In response, multimodal knowledge graphs (MMKGs) have emerged to serve the management and applications of different types of data. However, since the proposal of KG in 2012, there has not been a unified and standardized definition to describe KG, let alone MMKG. Based on previous research and experience, this paper has summarized the definition of KG through extensive investigation and explores the concept of MMKG. To provide a better illustration, this paper constructed a sample MMKG in the medical field based on an ontology and resource description framework (RDF). We use Neo4j for visualization and design a UI to extract node information. Finally, the shortcomings of the work were summarized, and future research directions were proposed.

Keywords: Multimodal Knowledge Graph · Ontology · Neo4j

1 Introduction

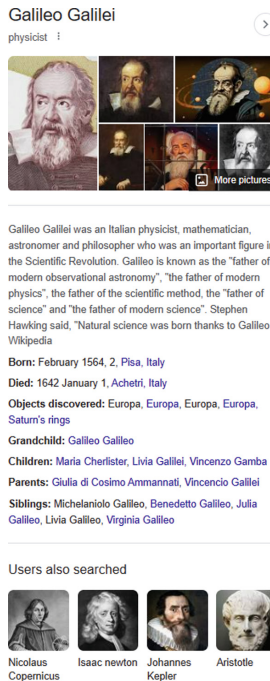
Knowledge is the crystallization of human understanding of the objective world in practice. As a form of structured knowledge, knowledge graphs (KGs) can be traced back to semantic nets proposed by Richens in 1956 [1, 2]. Later, expert systems such as MYCIN were proposed and became a research hotspot [3]. Expert systems were also considered the precursor of KG. As the builder of a search engine, Google is committed to understanding the words that users use. In other words, when users search, the words they enter refer to things that actually exist in the world, rather than just their surface meanings. Based on this idea, Google

This work was supported by the provincial and ministerial key project of the 14th Five-Year Scientific Research Plan of the State Language Commission in 2022 (ZDI145-58), and Xinjiang University PhD Start-up Fund Project (620320015).

attempted to establish relationships between real-world entities by building the KG. In 2012, it was integrated into the Google search engine, making it easier for users to access knowledge related to their search queries.

KG is a collection of information about real-world entities, including people, books, movies and many other types of things. For example, for a celebrity, relevant data such as their birthday and height are collected, and the person is linked to other closely related entities in the KG. More specifically, if a user wants to learn about astronomers, they may search for Galileo, as shown in Fig. 1. Based on the knowledge graph, the search result will directly display relevant information and show Galileo's scientific contributions. It can also help users discover other famous astronomers, such as Copernicus and Kepler. The goal of the KG is to move from an information engine to a knowledge engine.

The proposal of knowledge graphs has attracted widespread attention from academia and industry. As the knowledge graph continues to develop, it will become larger in scale and more content-rich. An increasing number of knowledge graphs are being created to support downstream applications such as knowledge management, search engines, intelligent question answering and recommendation systems. The research fields include: medical, archeology, e-commerce, catering, and economics [4-9].



Galileo Galilei
physicist

Galileo Galilei was an Italian physicist, mathematician, astronomer and philosopher who was an important figure in the Scientific Revolution. Galileo is known as the "father of modern observational astronomy", "the father of modern physics", the father of the scientific method, the "father of science" and "the father of modern science". Stephen Hawking said, "Natural science was born thanks to Galileo." [Wikipedia](#)

Born: February 1564, 2, Pisa, Italy
Died: 1642 January 1, Achetri, Italy
Objects discovered: Europa, Europa, Europa, Europa, Saturn's rings
Grandchild: Galileo Galileo
Children: Maria Cherlister, Livia Galilei, Vincenzo Gamba
Parents: Giulia di Cosimo Ammannati, Vincenzo Galilei
Siblings: Michelariolo Galileo, Benedetto Galileo, Julia Galileo, Livia Galileo, Virginia Galileo

Users also searched

Nicolaus Copernicus, Isaac Newton, Johannes Kepler, Aristotle

Fig. 1. Google search for Galileo

The concept of modal, similar to the concept of neural networks, was initially a biological concept. Humans have visual, auditory, tactile, and olfactory senses. Each different form of information can be referred to as a modal. In machine learning, it generally refers to different media of information, such as text, images, speech and videos. Multimodal refers to the combination of multiple different types of data. With the rapid development of the Internet, the explosive growth of information in different modalities has become a critical and challenging problem in terms of how to efficiently utilize these diverse types of information [24]. On the other hand, to overcome the limitations of a single mode in practical applications, the demand for machines to learn multimodal knowledge has also been increasing. For example, the image captioning task is one of the first tasks involving the combination of multimodal images and text. Machines need to automatically generate natural language descriptions of images, which requires more than the image understanding level provided by typical image recognition and object detection methods [12, 13]. Visual question answering is often seen as a visual Turing test, where the system needs to understand any form of natural language question (usually related to visual information in the image) and answer it in a natural way [17, 18].

However, as shown in Fig. 2(a), KGs mostly use pure symbolic text as objects, constructing a semantic network using triples. This approach limits the machines’ understanding and expression capabilities [11, 12]. If we only tell the machine about the description of “*dogs*”, it is difficult for the machine to understand the concept of “*dogs*”, which makes the application of KGs difficult. However, if we combine different modalities of information about dogs, such as pictures of dogs and the sound of barking, the image of “*dogs*” becomes vivid. In other words, if we want machines to truly gain intelligence, single-modal information alone is far from sufficient. Therefore, multimodal knowledge graph (MMKG), as shown in Fig. 2(b), has great help in achieving artificial intelligence. KGs are also urgently in need of multimodality.

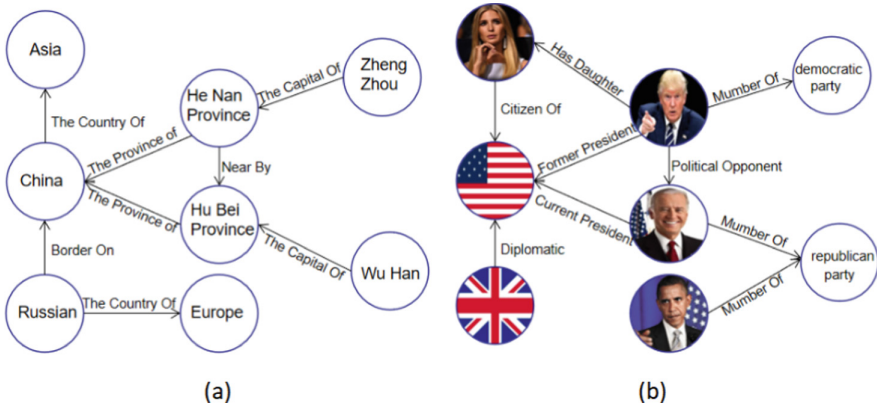


Fig. 2. (a) An example of unimodal KG (b) An example of a MMKG

In this context, the construction and application of MMKGs have become a research hotspot. However, there has been a thorny issue that has not been resolved, which is the definition of MMKG. Starting from the KG itself, this paper summarizes the definition of KG, explores the definition of MMKG, and provides an example MMKG in the medical field. The rest of this paper is organized as follows: Section 2 summarizes the definition of KG, and Sect. 3 explores the concept of MMKG. To illustrate the concept, Sect. 4 constructs an example MMKG in the medical field, and Sect. 5 provides a summary of the entire paper.

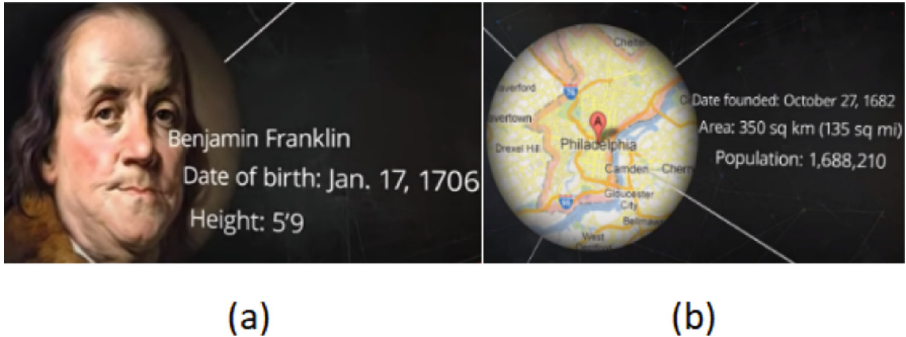


Fig. 3. (a) The introduction of Franklin (b) The introduction of Benjamin

2 Definition of KG

2.1 Description of the Problem

The definition of KG has been a longstanding topic of discussion among experts and scholars, but a consensus has not yet been reached [10–12, 19–23]. The root cause of this problem is that Google’s introduction to its KG blog did not mention the definition and related technical issues of KG, which has led to conflicting definitions and descriptions of KG in its development [1, 20]. For example, Paulheim et al. defined KG as a graph-based organization used to describe entities and their relationships in the real world [21]. This definition is too abstract and not sufficiently detailed to KG. Ehrlinger et al. defined KG as the acquisition of knowledge and integration into ontology, using a reasoning engine to deduce new knowledge [20]. The implication is that KG consists of two parts, knowledge and reasoning engine, which is also biased. Zheng et al. simply defined KG as representing entities with nodes and relationships with edges [4]. Most other papers mention the representation of KG, rather than its definition. For example, Ji et al. defined a knowledge graph as $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{F}\}$, where \mathcal{E} , \mathcal{R} and \mathcal{F} are sets of entities, relationships, and facts, respectively. Facts are represented as triples $\{h, r, t\} \in \mathcal{F}$ [23]. However, there is no distinction made between relationships

and attributes and no discussion of directivity between triples. As seen, even the representation of KG is difficult to have a unified standard [11]. This is very unfriendly for research in this field. Therefore, a unified and standard definition of KG is needed.

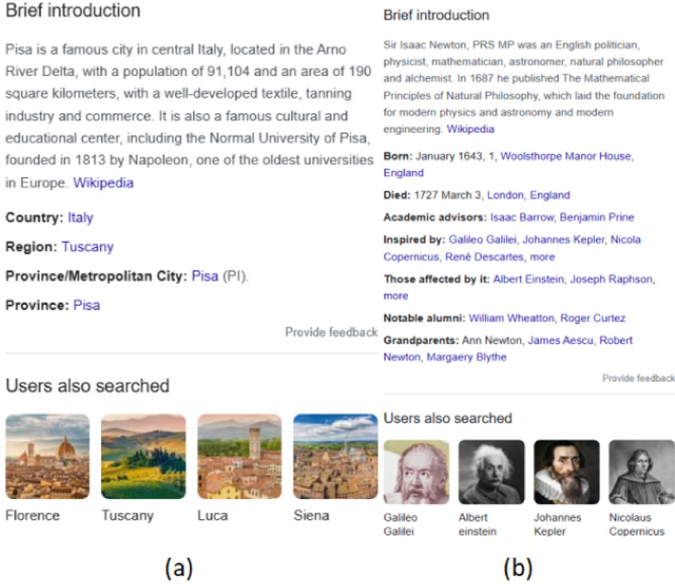


Fig. 4. (a) Introduction of Galileo’s birthplace, Pisa. (b) Introduction of Isaac Newton, a figure related to Galileo.

2.2 Inquiry into the Problem

To solve this thorny problem, must go back to the source and start with Google’s blog on KG. The blog provides a case of how KG is used for search, as shown in Fig. 1. We can see that the search result for Galileo consists of the following parts:

- The first part is Galileo’s name and classification: Galileo belongs to the category of physicists. We can view this classification as a part of the framework of ontology, and Galileo is an instance under this class.
- Then, are his images, which come from different sources such as BaiduPedia, StarWalk, Wikipedia [1].
- After the images, there is a section on Galileo’s personal information, including his life events, and contributions. The users could click the blue text and will have a page jump. The black text could not be clicked.

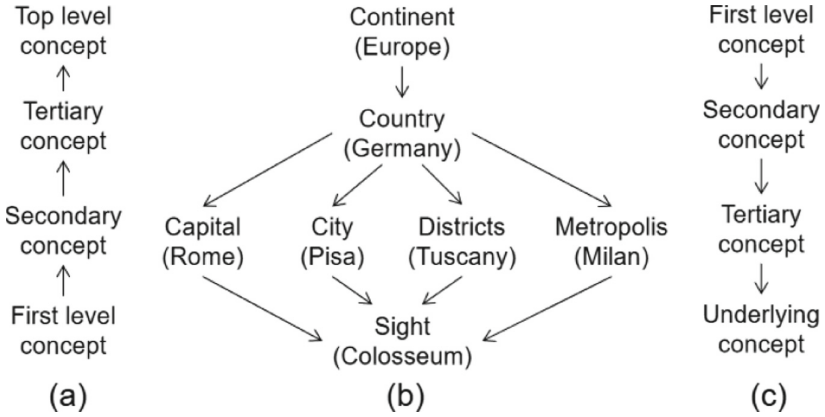


Fig. 5. (a) Inductive method for ontology construction (b) Ontology diagram (c) Taxonomic method for ontology construction

- At the bottom, there are related figures such as Copernicus, Newton. The text and images are integrated as a whole, and clicking on them can lead to their corresponding pages.

Therefore, we can see that the data in KG should include two types: resources and literals. Resources refer to resource links from different data sources. Literals can be understood as strings in programming languages, which do not have meaning in themselves. With the concept of “*trees*” in data structures, literals are similar to “*leaves*” with a degree of 0. Combined with another instance in the introduction, as shown in Fig. 3, resources exist in the form of entity nodes, and different entity nodes are connected through relationships, represented by white lines in the graph. Literals are usually considered internal information of entity nodes and are not connected to other nodes, which is called the property value. These two types of information can be described by triples. For example, “*Galileo - birth place - Pisa, Italy*”, “*Galileo - died in - January 8, 1642*”. The first triple was defined as entity-relationship-entity, and the other was defined as entity-property-property value. These two together form the basic components of a KG. One important point to note is the directivity of the entities in KG. Some literature mentions this issue, suggesting that KG should be defined as a directed graph structure [10, 11]. However, these studies has not provided a clear explanation on this issue: whether it is the directivity of the relationship between entities or the directivity between entities and property values, and whether this directivity refers to one-way or two-way, or multidirectional? This paper explains this issue: using the previous example, “*Galileo - died in - January 8, 1642*” is a reasonable expression, rather than “*January 8, 1642 - died in - Galileo*”. That is, in the triple of entity-property-property value, the node points to the property value, and the node’s property is only connected to that node, which is unidirectional. This matches the representation in Fig. 3. For the triple of entity-relationship-entity, such as “*Galileo - birth place - Pisa, Italy*”. An entity

is connected to many different entities, such as “Galileo”, which is related to many other figures, such as “Isaac Newton” and “Aristotle”; that is, the entity has multidirectionality. Furthermore, by clicking the Italy Pisa of birthplace information displayed in Fig. 1 and the recommended Isaac Newton below, as shown in Fig. 4, the recommended content displayed under the Pisa node is not related to Galileo, while Galileo appears in the recommended content below Newton. This indicates that the triple of “Galileo-birthplace-Italy, Pisa” is a one-way structure, while the triple of “Galileo-related person-Isaac Newton” is a two-way structure. That is, the relationship between nodes can be either one-way or two-way.

2.3 Knowledge Base, Ontology and RDF

Another point is the relationship and difference between knowledge graphs (KG) and knowledge bases. Many recent papers do not distinguish between these two concepts and treat KG and knowledge bases as equivalent [10,11]. They consider semantic networks, graph databases, and knowledge bases such as WordNet (1995), BabelNet (2010), Freebase (2008), DBpedia (2007), YAGO (2007), and WikiData (2014) as KG without explanation, which is obviously unreasonable [28–33]. One piece of evidence is that Johanna Wright, the product management director, mentioned in her introduction of KG that Google uses search engines to understand user search content and add some of this content to the knowledge base. This indicates that KG is a kind of knowledge base. However, other descriptions from Google employees suggest that these two concepts are not identical [1]. To this end, this paper explains that a knowledge base is a special database used for knowledge management. It is a collection of heterogeneous knowledge from multiple sources in a required field, including basic facts, rules, and other related information. A KG is a processed knowledge base that has a graph structure and contains structured and semistructured data. In addition to KG, two other frequently mentioned concepts are ontology and Resource Description Framework (RDF) [25,26]. RDF is a data model developed by W3C, which provides a uni-

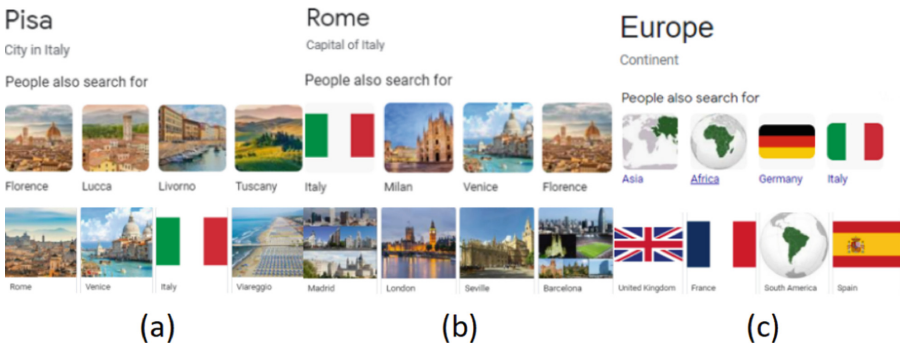


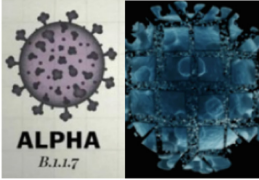
Fig. 6. (a) Search result of Pisa (b) Search result of Rome (c) Search result of Europe

fied standard for describing things and their relationships. RDF is composed of nodes and edges, where nodes represent specific entity resources or property values and edges represent relationships between entities or between entities and property values. RDF has constraints on each part of the SPO triple: “s” should be Internationalized Resource Identifiers (IRI) or blank node, “p” should be IRI and “o” could be IRI, resource or literals. However, RDF has a serious limitation in that it cannot distinguish between classes and objects. It also cannot define and describe class relationships and properties. In other words, RDF is mainly used to describe concrete things and lacks the ability to abstractly categorize and define groups of similar things. This clearly limits the expressive power of the model. Therefore, the assistance of an ontology is needed.

[TextCN data](#) [TextEN data](#) [Image data](#) [Voice data](#)
 Introduce: The data for Multi-Modal Knowledge Graph
 A website Used to store multimodal data to support downstream applications

Subject	Predicate	Object
Shortness of breath	involving_symptoms	Shallow breathing
Shortness of breath	clinic_department	Respiratory Medicine
Shortness of breath	belongTo	symptom
Functional dyspepsia	involving_symptoms	Loss of appetite
Functional dyspepsia	belongTo	disease
Functional dyspepsia	medical_college	GI Medicine
Unconsciousness	involve_checking	Arterial blood gas analysis
Unconsciousness	clinic	department Neurology

[Alpha1](#)
[Alpha2](#)
[Alpha3](#)
[Alpha4](#)
[Alpha5](#)
[Beta1](#)
[Beta2](#)
[Beta3](#)



ALPHA
B.1.1.7

[Chlamydia psittaci](#)
[COVID-19](#)
[Systolic bruit](#)
[Bronchial asthma](#)
[Macules](#)
[PLCH](#)
[Blepharoptosis](#)
[Bronchospasm](#)

▶ 0:04 / 0:04 ——— 🔊 ⋮

▶ 0:05 / 0:05 ——— 🔊 ⋮

▶ 0:16 / 0:16 ——— 🔊 ⋮

Fig. 7. Part of data of COMMKG-19

Ontology is a philosophical concept that involves dividing entities into basic categories and hierarchies. Ontology has a classification system and basic reasoning principles. The classification system defines the relationship between categories, providing the basis for reasoning. Some ontologies are widely used in the medical field, such as CIDO, GO, UberOn and DOID [34–37]. There are two main ways to build ontologies: the bottom-up inductive approach as shown in Fig. 5(a), and the top-down classification approach as shown in Fig. 5(c). Generally, the construction of open-domain KGs often uses the inductive method to classify features from underlying data due to the large amount of data involved. In contrary, domain-specific KGs often define classification categories before filling in the data. Google’s KG is full of the shadows of category in ontology. As previously mentioned, Galileo belongs to the category of physicists. As shown in Fig. 6(a), there is a comment line: “Pisa: The City of Italy”, which can be

viewed as a category label. In the recommended content below, we can see four recommended places, Florence, Lucca, Livorno and Tuscany. The category label of Florence, Lucca and Livorno is “*City in Italy*”, and the category label of Tuscany is “*Administrative districts of Italy*”. All of this recommended content belongs to places (cities or administrative regions) in Italy. In the additional recommendations, there are two items worth noting: one is “*Rome, the capital of Italy*”, and the other is “*Italy, Countries in Europe*”. When searching for Rome, as shown in Fig. 6(b), the related content includes Italy, Milan, Venice, and Florence. The corresponding category tags are countries in Europe and cities in Italy. In the additional recommendations, there are also Madrid (the capital of Spain) and London (the capital of the United Kingdom). It could be speculate that the recommended entities in the KG come from three categories: entities with the same label in the same category, with subcategory labels, and with parent category labels. Search for “*Europe*” to verify this assumption. As shown in Fig. 6(c), we obtained nodes with the same label: Asia and Africa. Subcategory nodes: Germany and Italy. The reason why there is no parent category entity is that “*Continent*” may be a top-level concept. The ontology based on this situation is shown in Fig. 5(b).

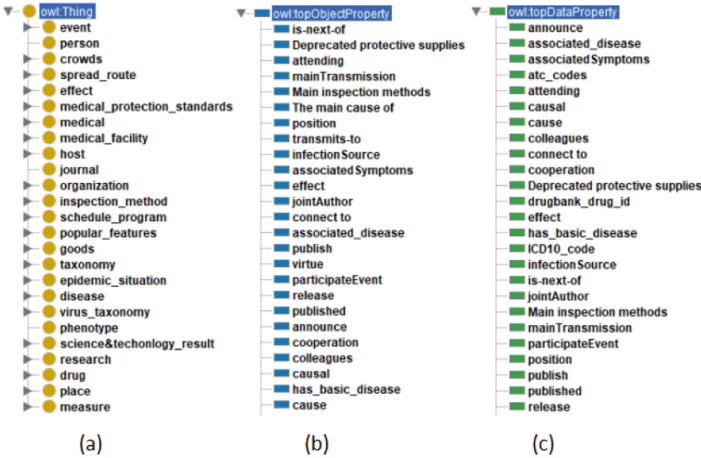


Fig. 8. (a) Some top-level concepts in the ontology (b) Some relationships in the ontology (c) Some properties in the ontology (c) Some properties in the ontology

2.4 Conclusion

In summary, this paper provides the definition of KG: KG is a kind of knowledge base composed of ontology and resource description framework, which can serve downstream applications. Its symbolic language is $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{P}, \mathcal{V}, \mathcal{T}_R, \mathcal{T}_P\}$, which is a set of elements and knowledge, where $\mathcal{E}, \mathcal{R}, \mathcal{P}, \mathcal{V}$ is a set of entities,

relationships, properties, property values. Knowledge $\mathcal{T}_{\mathcal{R}}$ is a set of triples of entity-relationship-entity, and $\mathcal{T}_{\mathcal{P}}$ is a set of triples of entity-property-property value. One piece of knowledge can be represented as $\mathcal{T}_{\mathcal{R}} = \{\mathcal{E}, \mathcal{R}, \mathcal{E}\}$ or $\mathcal{T}_{\mathcal{P}} = \{\mathcal{E}, \mathcal{P}, \mathcal{V}\}$. where \mathcal{R} and \mathcal{P} are directional, pointing from the head entity to the tail entity or property value. For example, “Zhengzhou belongs to Henan Province” can be expressed as $\mathcal{T}_{\mathcal{R}} = (\text{Zhengzhou}, \text{belongs to}, \text{Henan Province})$, and “Biden is 81 years old this year” can be expressed as $\mathcal{T}_{\mathcal{P}} = (\text{Biden}, \text{age}, 81)$.

3 Exploring the Concept of MMKG

Most literature researching MMKG does not mention the definition, and the definitions in some literature are too abstract [41,44]. Wang et al. directly introduced the RDF model into Richpedia and regarded it as a finite set of RDF triples [43]. Zhu et al. mentions that MMKG is a multimodal representation of part of the knowledge in KG [11]. In view of this phenomenon, it is necessary to summarize a unified and complete definition of MMKG.



Fig. 9. COMMKG-19 visualization by Neo4j

3.1 Multimodality of Knowledge Graphs

Some literature uses “CKG” to refer to knowledge graphs based solely on text modal [10]. This statement is unreasonable. The reason is that relevant researchers have ignored such a problem: Has the KG been multimodal since its inception? The answer is affirmative. The root cause of this problem is that Google does not mention the multimodal problem about KG in the relevant introduction. Although the concept of “multimodal” has been proposed for a long time, it was not until approximately 2015 that it received widespread attention

in the field of artificial intelligence, and most of the research was based on text and images [46]. As the key to Google’s search engine, the KG improves the performance of search engines in three ways: find the right thing, get the best summary and go deeper and broader. As the view in the blog, “*Language can be ambiguous-do you mean Taj Mahal the monument, or Taj Mahal the musician?*”, in order to provide better recommendations, KG has added image elements to relevant recommendations, such as Fig. 1 and Fig. 4. However, research based on KG, including construction and application, initially focused on text modal [47–52]. For example, Sören Auer built a KG for the exchange of academic information [49]. In the field of natural language processing (NLP), named entity recognition (NER) and relationship extraction(RE) work based on text modal has been greatly developed [50, 51]. With the proposal of the TransE model, knowledge representation learning(KRL) based on text information has become a major research hotspot [52, 53]. It is only in recent years that research on multimodality of KG has progressed. Examples include Liu et al.’s proposal of MMKG in 2019 and Wang et al.’s proposal of Richpedia in 2020 [43, 44].

It can be seen that the development of knowledge graphs is from multimodal to unimodal and then to multimodal. One of the main reasons for this is that: there is a lack of a unified understanding of the knowledge graph in academia and industry. This is why this paper explores the definition of a knowledge graph.

3.2 Comparison of KG and MMKG

Contrary to existing beliefs, based on the foregoing, this paper argues that MMKG should not be regarded as a generalization of KG; rather, KG is a special case of MMKG. In other words, KG is MMKG that contains only text modal information. Therefore, in terms of definition, MMKG and KG should conform to the same definition framework. Compared to the definition of KG, the definition of MMKG is broader. This paper defines MMKG as follows: MMKG is a kind of knowledge base that contains data in at least two different modals forms: text, voice, images, videos, etc. Follow the ontology and resource description framework, which can serve downstream applications. The symbolic language of MMKG is $\mathcal{G} = \{\mathcal{E}^M, \mathcal{R}^M, \mathcal{P}^M, \mathcal{V}^M, \mathcal{T}_{\mathcal{R}}^M, \mathcal{T}_{\mathcal{P}}^M\}$, where \mathcal{E}^M , \mathcal{R}^M , \mathcal{P}^M , and \mathcal{V}^M are a set of entities, relationships, properties, and property values and could be different modes. Knowledge $\mathcal{T}_{\mathcal{R}}^M$ is a set of triples of entity-relationship-entity with different modal, and $\mathcal{T}_{\mathcal{P}}^M$ is a set of triples of entity-property-property value with different modal. One piece of knowledge can be represented as $\mathcal{T}_{\mathcal{R}}^M = \{\mathcal{E}^M, \mathcal{R}^M, \mathcal{E}^M\}$ or $\mathcal{T}_{\mathcal{P}}^M = \{\mathcal{E}^M, \mathcal{P}^M, \mathcal{V}^M\}$, where \mathcal{R}^M and \mathcal{P}^M are directional, pointing from the head entity to the tail entity or property value.

The difference between KG and MMKG is mainly reflected in the application level, which is also the core issue of extensive research in academia. A major difficulty in researching MMKG is how to fuse the features of different modal data in a reasonable way to support downstream applications. Compared with the interaction between text modals in KG, MMKG needs to consider the features of different modal data. Current research focuses on supplementing text information with image information to improve the accuracy of downstream tasks. Sun

et al. designed a recommendation system based on the MMKG, which effectively alleviated the problems of cold start and data sparseness in the recommendation system [27]. Zheng et al. used doctor-patient dialogue and related examination pictures (CT, X-ray and ultrasound) to improve the accuracy of the diagnostic system for COVID-19 [39]. For KRL, the semantic information of unimodal limits the performance of the model. The introduction of modal data makes the performance of such models have more room for improvement. Wang et al. fused text and image modal features through the multihead self-attention mechanism to improve the accuracy of link prediction [55]. One thing to note is that image information should be as important as text information.

Table 1. Comparison of data between COKG-19 and COMMKG-19.

Model	Volume of Data						
	Concept	Property	Relationship	Entity	Triple	Speech	Image
COKG-19	505	393	82	26,282	32,352	/	/
COMMKG-19	512	397	84	26,432	60,039	268	2,700

4 Construction of MMKG

4.1 Two Different Ways to Build MMKG

At present, academia and industry generally use two different ways to construct MMKGs. One is to build MMKG using images as entity nodes. After that, the node information is enriched through the properties of the node, such as the size of the picture and the content of the image. This paper refers to this build as E-MMKG for short. Wang et al. built Richpedia following RDF. The text entity comes from Wikidata’s IRI. For image entities, collect images from Wikipedia and create corresponding IRIs in Richpedia. The result was a collection of 30,638 entities about cities, attractions, and celebrities. On average, a total of 99.2 images were retained for each entity. However, in Richpedia, the number of relationships between images is smaller and the ontology is simpler [43].

The other way is to build MMKG using the image as a property of the node, which this paper refers to as P-MMKG for short. Daniel et al. created ImageGraph, which contains 14,860 entities and 829,931 images. Its relationship structure is based on FB15K. For image data, more than 462 GB of image data was downloaded from different search engines. Corrupted, duplicate, and low-quality images are removed. In addition, triples in the header or tail entities that cannot be linked to the image data are filtered [54].

Compared with the construction method of E-MMKG, the construction of P-MMKG is simpler because the current attributes in the knowledge graph are not connected to other nodes, and there is no need to consider the relationship between images. The construction of E-MMKG often needs to consider the

relationship between images, such as similar or different. Although it enriches the amount of data in MMKG, it also increases the complexity of the build. In some fields where the relationship between images is not in high demand, the P-MMKG construction method is recommended. However, in some specific fields, such as in the Encyclopedia Knowledge Graph, illustrating the relationships between different species of animals through images (tigers and lions share a common ancestor), E-MMKG must be considered. These two different MMKG construction methods follow the ontology and RDF structure, which is consistent with the definition of MMKG in this paper.

4.2 Building Sample MMKG in the Medical Field

The potential of KG in the medical field is enormous and is considered the cornerstone for achieving smart healthcare. Some work based on KG in the medical field has made good progress [4, 39]. The outbreak of the COVID-19 virus in 2019 has had a profound impact on human life. Research based on the COVID-19 virus has been a hot topic in recent years. However, the shortcomings of these KGs are also very obvious: most medical KGs are based on textual data. A few MMKGs have limited types of image data, and these MMKGs do not consider speech data [39]. To provide a better illustration and facilitate better research by experts and scholars, this paper constructed a sample MMKG based on the COVID-19 virus, including textual, image and speech data.

Since there is no need to consider the relationship between images, this paper uses P-MMKG to construct the sample MMKG. The ontology and some of the textual data were referenced from COKG-19¹. COKG-19 is an open-source KG on COVID-19 primarily based on textual information jointly released by the AMiner team of the Department of Computer Science and Technology at Tsinghua University and the ZhikuAI team. The KG collected data from 8 COVID-19-related KGs that are open-source on OPENKG². Through various algorithms such as entity recognition, semantic matching and disambiguation, and knowledge fusion, the KG merged concepts with the same meaning, differentiated polysemous concepts, and supplemented and corrected them based on the opinions of relevant experts. In recent years, there have been some new variants of the COVID-19 virus. Therefore, this paper added some concepts, attributes, relationships, and instances to COKG.

For image data, a web crawler system is built to retrieve images related to entities from different search engines, which collect URL links to the top-ranking images of different search engines. Taking into account the cost of manual construction, the sample size is selected as 10% of the number of entities. To ensure the quality of the picture, we manually adjusted the size of some pictures and deleted low-quality pictures considering factors such as image size, clarity, and reliability. Filter out the most representative pictures as the property store of the node. It is worth mentioning that figurative pictures are chosen to convey

¹ https://covid-19.aminer.cn/kg/class/neurology_disease.

² <http://www.openkg.cn/>.

some non-visual concepts such as delirium. In the end, a total of 2700 pictures passed the screening, and some important nodes were assigned multiple images.

For speech data, the content of the dataset is mainly the clinical manifestations of partial symptoms. Through Text-To-Speech (TTS) technology, 268 audio files were generated using the open-source API of iFlytek³. We refer to the sample MMKG as COMMKG-19, in addition, COMMKG-19 additionally extracted English triples. The data pairs for COKG-19 and COMMKG-19 are shown in Table 1.

To store the above information and provide URL links, as shown in Fig. 7, this paper has established an open source website⁴. Protégé is an ontology editor developed by Stanford University, that is used to create and maintain ontologies and knowledge graphs. In this paper, Protégé was used to add, modify and supplement the ontology data of COKG-19, such as concepts, relationships and properties, as shown in Fig. 8.

The MMKG visualization was achieved by importing data into a Neo4j graph database by generating Turtle files, as shown in Fig. 9. In addition, to facilitate the extraction and utilization of MMKG data, a user interface was designed to retrieve node information, as shown in Fig. 10.

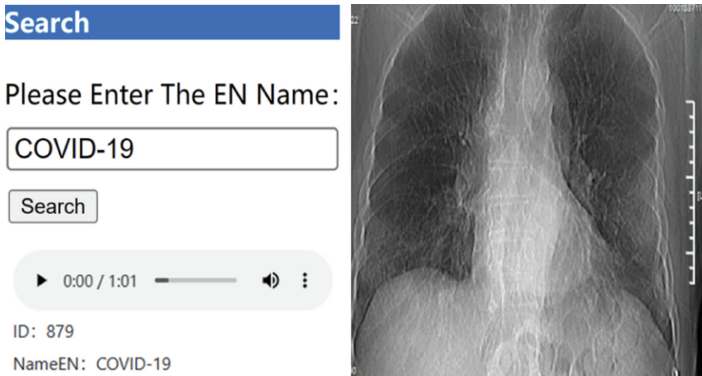


Fig. 10. User interface

5 Summary

In recent years, KG has made significant progress, and many MMKG-based studies have achieved remarkable advances. To promote a unified understanding of KG in the academic and industrial communities and to use the term “KG” more rigorously, this paper starts from the KG itself, conducts investigations

³ <https://www.xfyun.cn/>.

⁴ <https://xiangrui521.github.io/KnowledgeGraphData/>.

and research, summarizes previous work, proposes a definition of KG, explores the concept of MMKG, and provides a sample MMKG in the medical field.

The work presented in this paper has some limitations. First, the proposed definition of KG needs to be widely recognized and further refined by relevant researchers. Second, the MMKG sample constructed in the medical field has relatively few image and speech data, partly due to the high cost of manual work. Therefore, in future work, we will consider automated processing of speech and image data. In addition, video data have not been considered because there is currently limited research on video modal data, but video data often contain more information, which is an important aspect to consider. Due to article constraints, some of the content cannot be described in detail. We will focus on outlining the MMKG technical system to establish connections between different research fields and promote the development of the KG field in the future.

Acknowledgements. This work was supported by the provincial and ministerial key project of the 14th Five-Year Scientific Research Plan of the State Language Commission in 2022 (ZD1145-58), and this work was supported by the National Social Science Foundation of China (No. 22XYY048).

References

1. Singhal, A.: Introducing the Knowledge Graph: Things, not Strings, May 2012. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
2. Richens, R.H.: Preprogramming for mechanical translation. *Mech. Transl. Comput. Linguist.* **3**(1), 20–25 (1956)
3. Shortliffe, E. (ed.): *Computer-Based Medical Consultations: MYCIN*. Elsevier (2012)
4. Zheng, X., Xiao, Y., Song, W., et al.: COVID19-OBKG: an ontology-based knowledge graph and web service for COVID-19. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2456–2462. IEEE (2021)
5. Haase, P., Herzig, D.M., Kozlov, A., et al.: metaphactory: a platform for knowledge graph management. *Semantic Web* **10**(6), 1109–1125 (2019)
6. Do, P., Phan, T.H.V.: Developing a BERT based triple classification model using knowledge graph embedding for question answering system. *Appl. Intell.* **52**(1), 636–651 (2022)
7. Jiang, Z., Chi, C., Zhan, Y.: Research on medical question answering system based on knowledge graph. *IEEE Access* **9**, 21094–21101 (2021)
8. Fan, H., Zhong, Y., Zeng, G., et al.: Improving recommender system via knowledge graph based exploring user preference. *Appl. Intell.*, 1–13 (2022)
9. Gogleva, A., Polychronopoulos, D., Pfeifer, M., et al.: Knowledge graph-based recommendation framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer. *Nat. Commun.* **13**(1), 1667 (2022)
10. Peng, J., Hu, X., Huang, W., et al.: What is a multi-modal knowledge graph: a survey. *Big Data Res.*, 100380 (2023)
11. Zhu, X., Li, Z., Wang, X., et al.: Multimodal knowledge graph construction and application: a survey. *IEEE Trans. Knowl. Data Eng.* (2022)
12. Zhang, C., Yang, Z., He, X., et al.: Multimodal intelligence: representation learning, information fusion, and applications. *IEEE J. Sel. Top. Sig. Process.* **14**(3), 478–493 (2020)

13. He, X., Deng, L.: Deep learning for image-to-text generation: a technical overview. *IEEE Signal Process. Mag.* **34**(6), 109–116 (2017)
14. Zhang, Z., Xie, Y., Yang, L.: Photographic text-to-image synthesis with a hierarchically nested adversarial network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6199–6208 (2018)
15. Zhou, R., Jiang, C., Xu, Q.: A survey on generative adversarial network-based text-to-image synthesis. *Neurocomputing* **451**, 316–336 (2021)
16. Tan, H., Liu, X., Liu, M., et al.: KT-GAN: knowledge-transfer generative adversarial network for text-to-image synthesis. *IEEE Trans. Image Process.* **30**, 1275–1290 (2020)
17. Wu, Q., Teney, D., Wang, P., et al.: Visual question answering: a survey of methods and datasets. *Comput. Vis. Image Underst.* **163**, 21–40 (2017)
18. Wang, P., Wu, Q., Shen, C., et al.: FVQA: fact-based visual question answering. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(10), 2413–2427 (2017)
19. Wang, Q., Mao, Z., Wang, B., et al.: Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **29**(12), 2724–2743 (2017)
20. Ehrlinger, L., Wöß, W.: Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)* **48**(1–4), 2 (2016)
21. Paulheim, H.: Knowledge graph refinement: a survey of approaches and evaluation methods. *Semantic Web* **8**(3), 489–508 (2017)
22. Alani, H., et al. (eds.): *ISWC 2013. LNCS*, vol. 8219. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-41338-4>
23. Ji, S., Pan, S., Cambria, E., et al.: A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **33**(2), 494–514 (2021)
24. Ye, C., Zhou, G., Lu, J.: Survey on construction and application research for multimodal knowledge graphs. *Appl. Res. Comput.* **38**(12), 3535–3543 (2021)
25. Bunge, M.: *Treatise on Basic Philosophy: Ontology I: The Furniture of the World*. Springer, Dordrecht (1977). <https://doi.org/10.1007/978-94-010-9924-0>
26. Klyne, G.: *Resource description framework (RDF): concepts and abstract syntax* (2004). <http://www.w3.org/TR/rdf-concepts/>
27. Sun, R., Cao, X., Zhao, Y., et al.: Multi-modal knowledge graphs for recommender systems. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 1405–1414 (2020)
28. Miller, G.A.: WordNet: a lexical database for English. *Commun. ACM* **38**(11), 39–41 (1995)
29. Navigli, R., Ponzetto, S.P.: BabelNet: building a very large multilingual semantic network. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 216–225 (2010)
30. Bollacker, K., Evans, C., Paritosh, P., et al.: Freebase: a collaboratively created graph database for structuring human knowledge. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 1247–1250 (2008)
31. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: *DBpedia: a nucleus for a web of open data*. In: Aberer, K., et al. (eds.) *ASWC/ISWC - 2007. LNCS*, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_52
32. Suchanek, F.M., Kasneci, G., Weikum, G.: YAGO: a core of semantic knowledge. In: *Proceedings of the 16th International Conference on World Wide Web*, pp. 697–706 (2007)

33. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. *Commun. ACM* **57**(10), 78–85 (2014)
34. He, Y., Yu, H., Ong, E., et al.: CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. *Sci. Data* **7**(1), 181 (2020)
35. Ashburner, M., Ball, C.A., Blake, J.A., et al.: Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**(1), 25–29 (2000)
36. Mungall, C.J., Torniai, C., Gkoutos, G.V., et al.: Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* **13**(1), 1–20 (2012)
37. Schriml, L.M., Arze, C., Nadendla, S., et al.: Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* **40**(D1), D940–D946 (2012)
38. Bordes, A., Usunier, N., Garcia-Duran, A., et al.: Translating embeddings for modeling multi-relational data. In: *Advances in Neural Information Processing Systems*, vol. 26 (2013)
39. Zheng, W., Yan, L., Gou, C., et al.: Pay attention to doctor-patient dialogues: multimodal knowledge graph attention image-text embedding for COVID-19 diagnosis. *Inf. Fusion* **75**, 168–185 (2021)
40. Xiong, J., Liu, G., Liu, Y., et al.: Oracle bone inscriptions information processing based on multimodal knowledge graph. *Comput. Electric. Eng.* **92**, 107173 (2021)
41. Xu, G., Chen, H., Li, F.L., et al.: AliMe MKG: a multimodal knowledge graph for live-streaming e-commerce. In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 4808–4812 (2021)
42. Lei, Z., Haq, A.U., Zeb, A., et al.: Is the suggested food your desired?: Multimodal recipe recommendation with demand-based knowledge graph. *Expert Syst. Appl.* **186**, 115708 (2021)
43. Wang, M., Wang, H., Qi, G., et al.: Richpedia: a large-scale, comprehensive multimodal knowledge graph. *Big Data Res.* **22**, 100159 (2020)
44. Liu, Y., Li, H., Garcia-Duran, A., Niepert, M., Onoro-Rubio, D., Rosenblum, D.S.: MMKG: multi-modal knowledge graphs. In: Hitzler, P., et al. (eds.) *ESWC 2019*. LNCS, vol. 11503, pp. 459–474. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21348-0_30
45. Alberts, H., Huang, T., Deshpande, Y., et al.: VisualSem: a high-quality knowledge graph for vision and language. *arXiv preprint arXiv:2008.09150* (2020)
46. Baltrušaitis, T., Ahuja, C., Morency, L.-P.: Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(2), 423–443 (2018)
47. Wang, Z., et al.: XLORE: a large-scale English-Chinese bilingual knowledge graph. In: *ISWC (Posters & Demos)* (2013)
48. Wu, T., et al.: A survey of techniques for constructing Chinese knowledge graphs and their applications. *Sustainability* **10**(9), 3245 (2018)
49. Auer, S., et al.: Towards a knowledge graph for science. In: *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics* (2018)
50. Chiu, J.P.C., Nichols, E.: Named entity recognition with bidirectional LSTM-CNNs. *Trans. Assoc. Comput. Linguist.* **4**, 357–370 (2016)
51. Mintz, M., et al.: Distant supervision for relation extraction without labelled data. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (2009)
52. Bordes, A., et al.: Translating embeddings for modelling multi-relational data. In: *Advances in Neural Information Processing Systems*, vol. 26 (2013)
53. Lin, Y., et al.: Knowledge representation learning: a quantitative review. *arXiv preprint arXiv:1812.10901* (2018)

54. Oñoro-Rubio, D., et al.: Answering visual-relational queries in web-extracted knowledge graphs. arXiv preprint [arXiv:1709.02314](https://arxiv.org/abs/1709.02314) (2017)
55. Wang, E., et al.: Multimodal knowledge graphs representation learning via multi-headed self-attention. *Inf. Fusion* **88**, 78–85 (2022)