Panamerican Mathematical Journal
ISSN: 1064-9735
Vol 34 No. 3 (2024)

A Statistical Analysis of Knowledge Graph and its Applications

Reshma G. Nehe[1], Dr. Radhika Menon[2], Sagar D. Khedkar[3], Dr. N. S. Mujumdar[4],

Dr. D. Y. Patil Institute of Technology, reshmanehe6@gmail.com[1],
Dr. D. Y. Patil Institute of Technology, radhika.tharoor@gmail.com[2],
Symbiosis Skills and Professional University, sagardkhedkar11@gmail.com[3],
Rajarshi Shahu college of Engineering, mujum66@yahoo.com,[4]

Article History:

Received: *09-06-2024*

Revised: *16-07-2024*

Accepted: *23-08-2024*

Abstract:

In today's fast-paced world driven by big data and artificial intelligence, organizing and representing vast amounts of knowledge is crucial. Knowledge graphs are dynamic structures serve as repositories for real-world knowledge, presented in the form of interconnected graph data. With their ability to encapsulate complex information, knowledge graphs have swiftly captured the attention of both academia and industry alike. In this paper, we dive deep into understanding knowledge graphs. In first Section, we start by looking at their origins and evolution. In second section describes knowledge graphs, compares them with other graph forms, and explores how academic societies have incorporated them. In conclusion, we discuss the various applications of knowledge graphs in various domains and highlight the software tools that are facilitating their advancement. By providing a systematic overview of knowledge graphs, we aim to illuminate new avenues for research and foster advancements in this burgeoning field.

Keywords: Knowledge graph construction, Natural language processing, Data mining, Education.

1. Introduction

A Knowledge Graph (KG) is a structured knowledge presentation framework that collects and organizes information into a graph-like structure of nodes and edges. In KG, nodes represent entities such as objects, concepts or entities, while edges represent relationships or connections between these entities [16].

The concept of KG's has its roots in semantic networks and frame languages developed in artificial intelligence during the 1960s and 1970s. These early frameworks laid the groundwork for representing knowledge in a structured form. The term "knowledge graph" itself gained prominence after Google introduced its Knowledge Graph in 2012, which aimed to enhance its search engine's results with information gathered from a variety of sources, structured in a way that made sense semantically.

Since then, the evolution of knowledge graphs has been marked by the increasing adoption in various industries and the continuous development of technologies to construct, query, and maintain them. The rise of open standards like the Resource Description Framework (RDF) and Web Ontology Language (OWL) has also contributed to the widespread use of knowledge graphs, enabling a more unified approach to data representation and sharing across the web.

Today, knowledge graphs continue to evolve, incorporating advances in machine learning and natural language processing to become more dynamic and context-aware, further pushing the boundaries of how machines can understand and interact with human knowledge.

Recent research in KGs covers a diverse array of topics, showcasing the expanding significance of KGs across various domains. Notable areas of study include the utilization of Graph Neural Networks (GNNs) to embed KGs into low-dimensional spaces, enabling tasks like link prediction and entity classification. Additionally, ongoing research focuses on enhancing knowledge graph completion techniques, incorporating both local and global information to improve link prediction accuracy. Scalability and efficiency in KG construction, querying, and reasoning are also under scrutiny, with efforts directed towards distributed processing and compression methods to handle large-scale KGs more effectively. Moreover, researchers are exploring KGs for Explainable AI and Fairness, leveraging structured representations to enhance model interpretability and address biases. Furthermore, domain-specific applications of KGs, such as in healthcare and finance, are being explored, tailoring KGs to unique domain requirements for improved knowledge representation and analysis. Lastly, the integration of multimodal data sources into KGs is gaining traction, with research focusing on fusion techniques and applications across multimedia retrieval and recommendation systems. Overall, recent KG research spans a wide spectrum, from foundational algorithmic advancements to specialized applications, driving innovation and collaboration across interdisciplinary fields like machine learning and natural language processing. The concept of Linked Data [3] was published in 2009. The Semantic Web proposes to link different data series together so that they can be processed as one large global data graph. By 2014, approximately 1,000 datasets were linked together in the Linked Open Data cloud, most of the links between them connecting identical entities [18]. KGs can also help fight human trafficking. To help related organizations find traffickers and help victims, Szekely et al. [20] creates a wide field of human trafficking. They use sex trade ads that are constantly crawled from websites as a data source and combine this data from different sources using semantic technologies.

Software tools for KGs offer a diverse range of capabilities for creating, querying, and reasoning over KGs. Apache Jena provides robust APIs for RDF graph creation and SPARQL query processing, while Neo4j offers a powerful graph-based data model and Cypher query language. Stardog offers a comprehensive platform for RDF data modelling, SPARQL querying, and OWL reasoning, with integration capabilities. Amazon Neptune provides scalability and durability for cloud-based KGs, supporting property and RDF graph models. AllegroGraph is optimized for large-scale KG storage and querying, with RDF data modelling and SPARQL querying. Ontotext GraphDB specializes in semantic graph databases, offering SPARQL querying, inferencing, and full-text search. OpenLink Virtuoso provides a scalable triple store and SPARQL endpoint, supporting various RDF serialization formats and federated querying. Together, these tools offer a comprehensive suite of capabilities for managing and leveraging KGs across diverse domains and applications.

2. What is Knowledge graph:

Mathematically, KG can be defined as A directed labelled graph is a 4-tuple $G = (V, E, S, f)$, where V are Vertices, $E \subseteq V \times V$ are edges, S are labels, and $f: E \rightarrow S$ is a mapping function from vertices to symbols. People, companies, computers, and other objects can all function as vertices. An edge label represents the relationship of interest between the vertices, such as a friendship between two individuals, a customer relationship between a corporation and a person, a network link between two computers, and so on.

In other way, A Knowledge Graph is a comprehensive knowledge base designed to mirror the information flow within an organization [14], [17]. It adeptly stores both intricate structured data and unstructured knowledge, presenting them in the form of interconnected entities and their relationships. This expansive repository spans various topical domains, seamlessly acquiring and integrating new knowledge. Moreover, it facilitates the interrelation of diverse entities, enabling a holistic understanding of the interconnectedness inherent in the data [17], [22].

KGs are designed to encode rich and interrelated information about a specific domain or topic, facilating efficient information integration, retrieval, and inference. They allow complex relationships and semantics to be represented in a machine-readable format, which enables the collection and use of domain-specific information for various applications such as search, recommendation systems, query answering, and data analytics. KGs are commonly used in many fields such as healthcare, finance, e-commerce, natural language processing, and others, and play an important role in organizing and unlocking the potential of large-scale and heterogeneous data sources. Significance in AI and Data Management: In AI, knowledge graphs play a crucial role in enhancing machine understanding of complex data. They enable AI systems to reason about entities and their interrelations within a specific domain, which is vital for tasks like natural language processing, recommendation systems, and semantic search. In data management, knowledge graphs facilitate data integration, interoperability, and retrieval, making them essential for managing large-scale, heterogeneous data sources.

Construction of knowledge graph:

Data collection:

Compiling the fundamental information from multiple sources, such as documents, databases, and webpages. Identifying and differentiating entities (people, places, etc.) inside the gathered data model is known as entity identification.

Relationship extraction: Establishing the links between the entities that have been detected. Creating an ontology entail organizing the items and their connections using a clear framework. **Data storage**: Placing the knowledge graph in a database with specific functionality for managing graph data.

Querying: Searching, navigating, and exploring the network's connections using graph query languages.
Inference: Taking on complex assignments such as finding new connections or spotting discrepancies in the graph.

Following figure gives steps to construct knowledge graph.

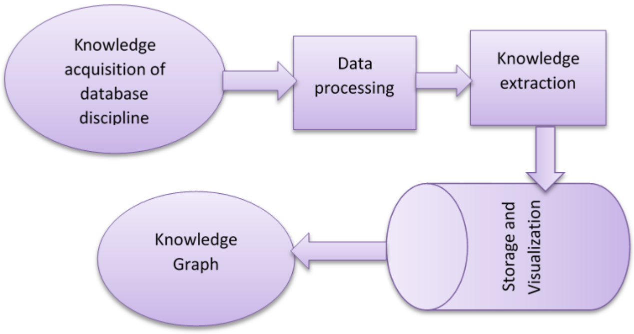


Figure: Construction of Knowledge Graph

Difference between KGs and other graph- based models.

This table highlights the key differences between KGs and other types of graphs, focusing on their structure, semantics, and applications. KGs are more structured and semantically rich, making them suitable for complex applications that require reasoning and inference, while other graphs are often used for analysing connections and network structures.

Aspect	Knowledge Graphs	Other Graphs
Semantic Meaning	Represent semantic knowledge with rich semantics using ontologies.	Lack a semantic layer; edges represent connections without specific meaning.
Structure	Structured with a predefined schema, well-defined entity types, and explicitly labelled relationships.	Unstructured, allowing arbitrary connections without specific semantics.
Multi-Relational Nature	Handle multi-relational data with various types of relationships.	Simpler, often representing only one type of relationship.
Focus on Entities and Attributes	Emphasize entities and their attributes, enriching the graph.	Focus more on connections, with less emphasis on entity attributes.
Reasoning and Inference	Support reasoning and inferencing based on transitive relationships and ontological axioms.	Inference is less common, e.g., social networks rarely involve reasoning.
Scalability and Complexity	Can be complex and challenging to scale due to diverse relationships and large-scale data.	Simpler and may be more scalable but lack the richness of KGs.
Use of Standards	Adhere to standards like RDF and OWL.	May not follow specific standards.
Applications	Power applications like question answering, recommendation systems, and semantic search.	Serve different purposes like social analysis and network centrality.

Knowledge Graphs and their correspondence with academic communities:

Knowledge Extraction: Knowledge extraction involves populating Knowledge Graphs by extracting structured information from unstructured data sources such as text and images. Natural Language Processing researchers focus on developing techniques for information extraction, entity linking, and relation extraction to extract meaningful knowledge from text data. Additionally, techniques from Data Mining contribute to automating the extraction of facts and relationships from various data sources. Information Retrieval researchers explore methods for retrieving relevant knowledge from large corpora, enabling the extraction of valuable insights from vast amounts of data. Together, these interdisciplinary efforts drive advancements in knowledge extraction, facilitating the creation of comprehensive and enriched KG that capture diverse domains of knowledge. [7]

Representation Learning: Representation learning plays a pivotal role in creating meaningful embeddings for KG entities and relations, aiming to capture semantic similarities and contextual information within the graph structure. Machine Learning (ML) researchers are at the forefront of developing representation learning techniques, such as graph neural networks, which are tailored to the unique characteristics of KGs. These techniques leverage graph structure and node features to learn embedding's that encode rich semantic information about entities and relationships. Concurrently, researchers in the Semantic Web community delve into ontological embeddings and semantic similarity measures, exploring how to embed entities and concepts into vector spaces that preserve their semantic relationships and properties. Moreover, representation learning in KGs aligns closely with research in Graph Theory, as it involves analysing and leveraging the structural properties of graphs to learn informative embeddings. By integrating insights from machine learning, semantic web, and graph theory, representation learning techniques contribute to the creation of expressive and semantically rich embeddings for KGs, facilitating various downstream tasks such as link prediction, entity classification, and recommendation systems.

Reasoning and Inference: Reasoning is a fundamental process in KGs that involves making logical deductions based on existing facts within the graph, enabling the system to answer complex queries and infer new knowledge. In the realm of AI, researchers are engaged in various forms of reasoning, including rule-based reasoning, probabilistic reasoning, and semantic reasoning, to enhance the cognitive capabilities of KG-based systems. Additionally, researchers in the field of Databases concentrate on developing efficient query answering techniques and scalable reasoning algorithms to handle the growing volume and complexity of KGs. Moreover, contributions from the Logic and Knowledge Representation communities play a crucial role in advancing reasoning techniques, drawing upon formal logic and knowledge representation frameworks to develop sound and effective reasoning methods. Together, these interdisciplinary efforts drive innovation in reasoning approaches, empowering KG-based systems to perform sophisticated inference tasks and derive meaningful insights from vast repositories of structured knowledge.

Fusion and Integration: Fusion plays a crucial role in integrating heterogeneous data sources into a coherent KG, combining information from structured databases, textual sources, and other modalities to enrich the graph with diverse and complementary knowledge. In database research, data integration and fusion techniques are extensively studied to address the challenges of integrating disparate data sources and ensuring data consistency and coherence within the KG. Similarly, researchers in the

Semantic Web community delve into ontology alignment, schema matching, and data fusion techniques to facilitate the seamless integration of data from different domains and sources. Furthermore, fusion techniques align closely with research in Multimodal Learning, where the integration of textual, visual, and other data modalities is explored to enhance learning and understanding from heterogeneous data. By leveraging insights from databases, the Semantic Web, and multimodal learning, fusion techniques enable the creation of comprehensive and enriched KGs that capture diverse perspectives and knowledge representations, fostering interdisciplinary collaboration and driving advancements in data integration and knowledge discovery.

Quality Assessment and Cleaning: Quality Assessment and Cleaning are essential processes in KG management, focusing on identifying inconsistencies, resolving contradictions, and enhancing the reliability of the data contained within the graph. In the realm of Data Management, researchers specializing in data quality and data cleaning techniques play a pivotal role in developing methodologies and algorithms for assessing and improving the quality of KGs. Similarly, in the field of Web Science, there is a concerted effort to ensure the quality of KGs, aligning with research on web data quality and integrity [29]. Furthermore, researchers in Information Systems delve into topics such as data provenance and trustworthiness, aiming to establish mechanisms for tracking the origin and reliability of data within KGs. By addressing these challenges and leveraging insights from data management, web science, and information systems, KG quality assessment and cleaning efforts contribute to maintaining the integrity and usefulness of KGs for diverse applications and domains.

3. Applications across various domains:

Natural Language Processing (NLP): KGs serve as a foundational framework for advancing NLP by capturing and representing rich semantic relationships among entities and concepts in a structured format. In NLP applications, KGs play a crucial role in enhancing language understanding and enabling more sophisticated text analysis tasks. KGs facilitate entity linking and disambiguation, allowing NLP systems to accurately identify and resolve references to real-world entities mentioned in text. Furthermore, KGs provide contextual information and background knowledge that can enrich language understanding by capturing the relationships between entities, their attributes, and the context in which they appear. KGs enable semantic search and question answering systems to retrieve relevant information from vast text corpora by leveraging the structured knowledge encoded in the graph. Bordes et al. [6], [5] introduced an embedding-based framework for this task. Additionally, KGs support information extraction and summarization tasks by providing a structured representation of textual information, which aids in extracting key facts and generating concise summaries. Through their ability to integrate diverse sources of knowledge and support reasoning over complex relationships, KGs significantly enhance the capabilities of NLP systems, enabling them to perform more sophisticated language understanding tasks with higher accuracy and efficiency. KGs enhance QA systems by providing structured knowledge for answering user queries [26]. For instance, the Google Knowledge Graph powers Google's search results with factual information.

Named Entity Recognition (NER): KGs play a pivotal role in enhancing (NER) systems by providing contextual knowledge and semantic relationships between entities. In NER, KGs serve as a valuable source of background information for disambiguating named entities mentioned in text. By leveraging the structured information encoded in the KG, NER systems can accurately identify and classify named entities into predefined categories such as persons, organizations, locations, dates, and more KGs enable NER systems to resolve ambiguous references by considering the relationships between entities and their attributes. For example, in the context of a news article mentioning "Apple," a KG can help determine whether it refers to the technology company, the fruit, or another entity by analysing the connections between "Apple," its associated attributes (e.g., headquarters, products), and related entities (e.g., Steve Jobs, iPhone). Additionally, KGs facilitate entity normalization by linking recognized entities to unique identifiers or URIs within the graph, enabling consistent representation and integration of entity references across different documents and datasets. Overall, KGs significantly enhance the accuracy and robustness of NER systems by providing a rich source significantly enhance the accuracy and robustness of NER systems by providing a rich source of contextual knowledge and semantic relationships for entity recognition and disambiguation. For more information one can refer [25].

Data Mining: KGs offer several applications in the field of Data Mining, enriching the process of extracting valuable insights and patterns from large and complex datasets

Pattern Discovery and Association mining: KGs provide a structured representation of relationships between entities attributes and concepts, facilitating the discovery of interesting patterns and associations within the data. By analysing the connections between entities in the graph, data mining algorithms can uncover hidden relationships and correlations, leading to insights that may not be apparent from individual data points alone.

Anomaly Detection: KGs can be leveraged for anomaly detection by modelling the normal behaviour of entities and detecting deviations from expected patterns. By analysing the connections and properties of entities within the graph, data mining techniques can identify unusual or suspicious patterns that may indicate fraudulent activities, system failures, or other anomalies.

Semantic Data Integration: KGs serve as a unifying framework for integrating heterogeneous data from diverse sources, including structured databases, text corpora, and sensor networks. By mapping disparate data sources to a common ontology and representing them as nodes and edges in the graph, data mining algorithms can perform integrated analysis across different data modalities, enabling comprehensive insights and understanding.

Predictive modelling: KGs provide a rich source of contextual knowledge and background information that can enhance predictive modelling tasks. By incorporating relevant information from the graph, such as entity attributes, relationships, and historical patterns, data mining algorithms can build more accurate and robust predictive models for various applications, including forecasting, classification, and regression.

Graph based clustering and classification: KGs can be directly used as input for graph-based clustering and classification algorithms, which group similar entities together based on their structural and semantic similarities within the graph. By partitioning the graph into clusters or predicting labels for nodes based on their connectivity patterns and attributes, data mining techniques can uncover meaningful structures and groupings in the data, facilitating exploratory analysis and knowledge discovery.

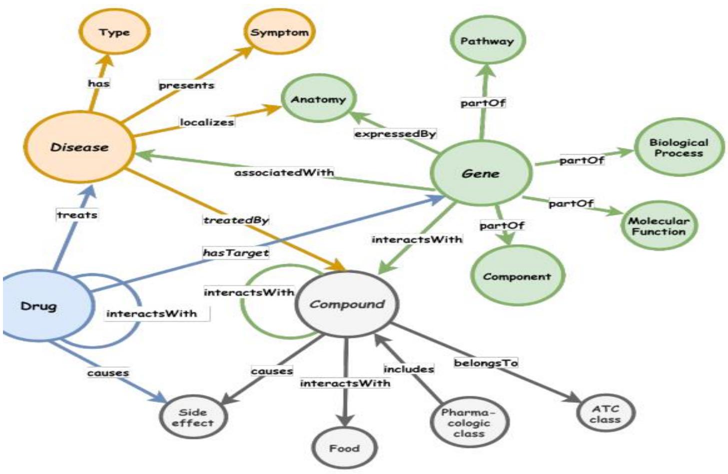
Linked Data: KGs adhere to linked data principles, allowing seamless integration of distributed information on the web [4].

Schema Matching: KGs aid in mapping data from different schemas, crucial for data integration

Healthcare and Life Sciences: KGs offer invaluable contributions to the domains of healthcare and life sciences by organizing, integrating, and analysing vast amounts of heterogeneous data, ranging from patient records and clinical trials to molecular structures and genetic information. In healthcare, KGs empower clinicians with comprehensive patient profiles, enabling personalized treatment plans and predictive analytics for disease management. They facilitate interoperability among disparate healthcare systems, fostering seamless data exchange and collaboration across healthcare providers. Furthermore, KGs drive biomedical research by aggregating data from various sources, facilitating hypothesis generation, and accelerating drug discovery and development processes [8]. KGs track adverse drug reactions and interactions, also assist doctors by providing evidence-based recommendations and insights.

In life sciences, KGs play a pivotal role in understanding complex biological systems, elucidating genotype-phenotype relationships, and uncovering novel biomarkers and therapeutic targets. By leveraging semantic technologies and advanced analytics, KGs enhance data-driven decision-making, foster innovation, and ultimately contribute to improving patient outcomes and advancing scientific knowledge in healthcare and life sciences. For KG applications in biomedical and health sciences one can refer [27, 28].

In order to get important insights, a biomedical knowledge graph can be used to depict the intricate relationships between biological elements. Here's a schematic illustration of how that may appear:



A General Biomedical knowledge graph (Source: Healthcare Knowledge Graph Construction)

E-Commerce and Retail: KGs serve as versatile tools across various industries, revolutionizing processes and decision-making. In financial services, they aid in fraud detection by linking suspicious entities and assessing credit risk through borrower-lender relationships, while also assisting in portfolio management by identifying asset correlations. In smart cities and urban planning, KGs optimize transportation and energy management by modelling traffic flow and energy consumption patterns, as well as enhancing urban resilience through disaster response planning. Meanwhile, in media and entertainment, KGs enable personalized content recommendations, event discovery, and copyright management. Overall, KGs significantly impact information retrieval, decision-making, and knowledge discovery, with their widespread adoption benefiting academia and industry alike.

Education: Knowledge graphs are increasingly being recognized as a valuable tool in the field of education. They can enhance knowledge management processes at universities and other educational institutions by facilitating the individualization of education. This allows for a more personalized learning experience, where students can select courses based on their preferences and career goals [13]. In essence, knowledge graphs organize information in a way that is both accessible and meaningful, making it easier for students and educators to find and use the data they need. The application of knowledge graphs in education can lead to improved outcomes, such as better understanding of complex subjects and more efficient research [9].For instance, in water conservancy education, knowledge graphs have been used to manage and organize educational resources effectively, aligning them with self-directed learning objectives [2]. Similarly, a systematic literature review [10] has shown that knowledge graph construction methodologies can be applied across various domains in education to enhance learning and teaching.

Knowledge graphs play a crucial role in organizing educational content, improving learning experiences, and supporting research in the field of education. These applications demonstrate the potential of knowledge graphs to transform educational systems by providing a structured, interconnected web of knowledge that supports learning and teaching activities. KGs can recommend personalized learning paths based on a student's progress, interests, and prerequisites. It also helps learners discover relevant educational resources efficiently. Identifying prerequisites for courses using KGs. Analysing student behaviour and performance within educational contexts. KGs enhance learning by providing structured and interconnected knowledge representations [21].

Some limitations are shown in relevant studies of application of knowledge graph in education

a) There is a lack of information on the knowledge graph construction process and methods, making it difficult to duplicate or adapt proposed solutions to unique settings or domains.

b) It is challenging to apply knowledge graph techniques to different subject areas. Because educational domains differ in terms of content, structure, pedagogical approaches, and learning objectives, the techniques and strategies used to construct knowledge graphs and extract relevant information may be influenced by the unique characteristics and requirements of each domain.

c) The resources used to create knowledge graphs are limited in size and scope. The data sources are frequently limited to specific topic areas, educational levels, or institutions, which limits the generalizability of the resulting knowledge graphs [24].

Cybersecurity: KGs have numerous applications in cybersecurity, leveraging their ability to represent complex relationships and contextual information. Here are some applications [19]:

Threat Intelligence Management: KGs can aggregate and organize threat intelligence data from various sources such as feeds, reports, and threat databases. By linking entities like attackers, malware, vulnerabilities, and tactics, techniques, and procedures (TTPs), KGs enable better understanding of the threat landscape.

Incident Response and Forensics: During incident response, KGs can help investigators quickly understand the scope and impact of an incident by visualizing relationships between affected systems, users, IP addresses, and other relevant entities. This aids in identifying the attack vector, understanding the attack chain, and attributing the incident to threat actors.

Vulnerability Management: KGs can integrate data from vulnerability scanners, asset inventories, and patch management systems to prioritize vulnerabilities based on their exploitability, impact, and affected assets. By mapping vulnerabilities to affected systems and their dependencies, organizations can make informed decisions regarding patching and remediation.

Identity and Access Management (IAM): KGs can model user roles, permissions, group memberships, and access policies within an organization's IT infrastructure. This enables organizations to detect anomalous access patterns, enforce least privilege principles, and identify potential insider threats or unauthorized access.

Network Security and Threat Detection: KGs can represent network topologies, configurations, traffic flows, and security policies. By analysing network data and correlating it with contextual information from the KG, organizations can detect suspicious activities, intrusions, or anomalies indicative of a security breach.

Security Automation and Orchestration: KGs can serve as a foundation for security automation by providing a unified representation of security-related data and processes. This enables the development of automated workflows for incident response, threat hunting, policy enforcement, and other security operations.

Compliance Management: KGs can facilitate compliance management by mapping regulatory requirements to specific controls, assets, and processes within an organization. This enables organizations to assess their compliance posture, identify gaps, and prioritize remediation efforts.

Threat Hunting and Proactive Defence: KGs enable security analysts to conduct proactive threat hunting by exploring relationships between seemingly unrelated entities such as indicators of compromise (IOCs), infrastructure elements, and threat actor behaviours. This helps identify emerging threats and anticipate potential attack vectors.

Overall, knowledge graphs play a crucial role in enhancing situational awareness, decision-making, and automation across various cybersecurity domains.Cybersecurity knowledge graph construction and for its applications one can refer [30].

Geoscience: In the big data era, KGs are pivotal in geoscience for structuring knowledge through maps, texts, and numerical data, leading to a unified geoscience knowledge model. KGs deepen big data analysis, enhance geological time scale accuracy, aid in creating intelligent maps, and facilitate knowledge evolution and reasoning in geoscience. These tools are more than mere information organizers; they're innovation platforms that merge geoscience with information and data science, propelling new research and significant theoretical advancements. The diverse uses of geoscience knowledge graphs have the potential to foster the convergence of data science, information science, and geoscience and hasten the advancement of these fields. By fusing the statistical and physical characterizations of geo-big data, knowledge-driven spatiotemporal analysis of big data can help

Panamerican Mathematical Journal
ISSN: 1064-9735
Vol 34 No. 3 (2024)

implement more accurate geo-big data analysis and enhance the comprehensive analysis. By utilizing the current geoscience information base and knowledge engine, it is possible to further the study of the geoscience knowledge system, comprehend the features of its evolution, and even uncover new geoscience knowledge through discoveries and innovations. Combining geoscience knowledge with earth system models may accelerate exploration, and integrating cartography knowledge with geoscience knowledge can encourage the intelligent and autonomous growth of map making. [11,23]

To create a geoscience knowledge graph (KG), begin by collecting georeferenced data from various sources, including rock parameters and satellite imagery. Next, use natural language processing to discover important concepts and correlations from this data and extract geological information. After defining these ideas and their relationships with an organized representation model, build the KG by constructing nodes and edges to represent geological things and their relationships. Create a thorough geoscience knowledge graph by integrating various knowledge elements and employing federated approaches for both quality and currency. Lastly, use the KG for detailed analysis to produce intelligent maps and improve geological time scales. A practical foundation for KG creation is provided via a case study on the extraction of data from mining exploration, which sheds light on the formalization of geological knowledge representation.

Applications of KGs is summarised with the help of following graphical representation briefly.

4. Software tools used for KG:

Apache Jena: It is an open-source Java framework for building semantic web and linked data applications. It provides APIs for creating and querying RDF (Resource Description Framework) graphs, as well as supporting various RDF serialization formats and SPARQL query processing.

Neo4j: It is a popular graph database management system that is widely used for building and querying Knowledge Graphs. It offers a powerful graph-based data model and a query language called Cypher, which allows for expressive graph pattern matching and traversal.

Stardog: It is a Knowledge Graph platform that provides capabilities for creating, querying, and reasoning over large-scale KGs. It supports RDF data modelling, SPARQL querying, and OWL (Web Ontology Language) reasoning, as well as integration with other data sources and systems.

Amazon Neptune: It is a fully managed graph database service offered by Amazon Web Services (AWS). It supports both property graph and RDF graph models, providing high availability, durability, and scalability for building and querying Knowledge Graphs in the cloud.

AllegroGraph: It is a graph database platform optimized for storing and querying large-scale Knowledge Graphs. It supports RDF data modelling, SPARQL querying, and reasoning capabilities based on RDFS (RDF Schema) and OWL.

Ontotext GraphDB: Ontotext GraphDB is a semantic graph database that specializes in managing and querying RDF data. It offers features such as SPARQL querying, inferencing, and full-text search, as well as support for ontology management and versioning.

OpenLink Virtuoso: OpenLink Virtuoso is a scalable and high-performance triple store and SPARQL endpoint that can be used for building and querying Knowledge Graphs. It supports various RDF serialization formats, SPARQL query optimization, and federated querying over distributed data sources.

5. Research Trends and Prospects:

A few issues should be investigated for the domain knowledge graph construction research [9]. **Knowledge representation expansion**: The relational triple is the primary form of knowledge expression; it can be expanded to a multicomponent form to communicate a variety of facts. Is it possible to extend characteristics and relationships for complicated unstructured situations that a basic inclusion relationship cannot adequately explain, such the corresponding relationship between major and school in the education industry?

Multiple representation modes for knowledge: The network has a vast amount of information resources. They feature images, videos, and other content in addition to text. Perhaps images and videos convey information more clearly than words. Thus, an essential difficulty is how to incorporate these nontextual resources into the knowledge representation system.

Knowledge representation automatic learning: The majority of techniques for learning knowledge representation are transferable to a broad knowledge graph. For the automatic extraction of domain knowledge with complicated information, it is not very qualified. Thus, automatic learning of domain knowledge is an open problem.

Knowledge fusion: When it comes to integrating knowledge into an industry context, domain knowledge has a more intricate structure than general knowledge, necessitating greater data testing and improved algorithms.

Data gathering: A knowledge graph's development necessitates a large amount of data. A comprehensive education knowledge graph can help students gather information and teachers with course planning in certain specialized fields, including education. The creation of an education knowledge graph is supported by an adequate amount of data. In the education sector, data collecting is still ongoing, though. Thus, the development of domain knowledge graphs will be a research topic involving data collection and analysis platforms.

Dynamic update: Certain types of knowledge, including major courses taken each year, school grade, or major in the education area, are not always constant in the domain knowledge graph. Thus, one major topic of research is how to quickly update and actualize dynamic changes in knowledge.

6. Conclusion

This review paper has provided a comprehensive examination of knowledge graphs, delving into their definition, applications, recent research trends, and the software utilized in their development and utilization. Knowledge graphs, as intricate representations of interconnected concepts, entities, and relationships, have emerged as powerful tools across various domains, from semantic web technologies to artificial intelligence applications. Differentiating from traditional graphs, knowledge graphs encapsulate not only the structural connections between nodes but also the rich semantic information associated with each entity and relationship.

This semantic richness enables more nuanced representations of real-world knowledge, fostering deeper insights and enabling more sophisticated reasoning capabilities. Through our exploration, it has become evident that knowledge graphs facilitate the organization and retrieval of information in a structured and semantically meaningful manner, enabling more effective data management, reasoning, and decision-making processes. Their versatility is showcased through diverse applications such as natural language processing, recommendation systems, and biomedical informatics.

Moreover, recent advancements in knowledge graph research have focused on enhancing their scalability, interpretability, and interoperability, addressing challenges such as entity resolution, link prediction, and knowledge fusion. Emerging techniques such as deep learning and probabilistic reasoning have been integrated to further improve knowledge graph capabilities, paving the way for more intelligent and adaptive systems.

In terms of software tools, a myriad of frameworks and platforms have been developed to facilitate knowledge graph construction, querying, and analysis, including popular options like Neo4j, RDFLib, and Apache Jena. Each tool offers distinct features and functionalities tailored to different use cases, contributing to the growing ecosystem of knowledge graph technologies.

In essence, knowledge graphs represent a paradigm shift in how we conceptualize and leverage data, offering a powerful means to organize, integrate, and harness the ever-expanding repositories of human knowledge. As research in this field continues to evolve, we anticipate even more transformative applications and methodologies to emerge, further solidifying the pivotal role of knowledge graphs in shaping the future of information management and artificial intelligence.

7. Refrences

[1] Abu-Salih, B. A systematic literature review of knowledge graph construction and https://doi.org/10.1016/j.heliyon.2024.e25383

[2] Application of Knowledge Graph in Water Conservancy Education … - MDPI. https://www.mdpi.com/2079-9292/11/23/3913.

[3] Bizer, C., Heath, T., Berners-Lee, T. (2011). Linked data: The story so far. In: Semantic services, interoperability and web applications: emerging concepts, pp. 205–227. IGI Global.

[4] Blumauer, A. (2014). From Taxonomies over Ontologies to Knowledge Graphs The Semantic Puzzle, https://semantic-web.com/2014/07/15/from-taxonomies-over-ontologies-to-knowl- edge-graphs [accessed: 17.02.2019].

[5] Bordes, A., Chopra, S., & Weston, J. (2014). Question answering with subgraph embeddings. In Proc. Conf. Empirical Methods Natural Language Process., pp. 615–620.

[6] Bordes, A., Weston, J., & Usunier, N. (2014). Open question answering with weakly supervised embedding models. In Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases, pp. 165–180.

[7] Jaradeh, M. Y., Singh, K., Stocker, M., Both, A., & Auer, S. (2023). Information extraction pipelines for knowledge graphs, Knowledge and Information Systems, 65, 1989–2016. https://doi.org/10.1007/s10115-022-01826-x.

[8] Jiang, Z., Chi, C., & Zhan, Y. (2021). Research on Medical Question Answering System Based on Knowledge Graph, pp. 21094-21101.

[9] Jinjiao Lin, Y., Zhao, Y., Huang, W., Liu, C., & Pu, H. (2021). Domain knowledge graph-based research progress of knowledge representation, Neural Computing and Applications, 33, 681–69. https://doi.org/10.1007/s00521-020-05057-5.

[10] Kejriwal, M. (2022). Knowledge Graphs: A Practical Review of the Research Landscape. Information, 13(4), 161.

[11] Knowledge graph construction and application in geosciences: A review … https://eartharxiv.org/repository/object/2312/download/4746.

[12] Mariia Rizun, Knowledge Graph Application in Education: a Literature Review, Acta universities lodziensis. folia oeconomica (aug 2019), Vol. 3, no. 342, pp. 7 – 19.

[13] Mariia Rizun, Knowledge Graph Application in Education: a Literature Review, Folia Oeconomica Acta Universitatis Lodziensis, 3(342), 2019. DOI: http://dx.doi.org/10.18778/0208-6018.342.01.

[14] Masuch, L., Muszynski, H., & Raethlein, B. (2014). Enterprise Knowledge Graph, www.slideshare.net/BenjaminRaethlein/enterprise-knowledge-graph–56858153 [accessed: 11.06.2018].

[15] Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs, Proceedings of the Institute of Electrical and Electronics Engineers, 104(1), 11–33. https://arxiv.org/abs/1503.00759 [accessed: 27.05.2018].

[16] Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation z www.semantic-web-journal.net/system/files/swj1167. pdf [accessed: 27.05.2018].

[17] Pirrò, G. (2015). Explaining and suggesting relatedness in knowledge graphs. In: International Semantic Web Conference. In: M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan and S. Staab (eds.), The Semantic Web – ISWC 2015. ISWC 2015. Lecture Notes in Computer Science, vol. 9366. Springer, Cham, http://disi.unitn.it/~pavel/OM/articles/93660561.pdf [accessed: 17.02.2019].

[18] Schmachtenberg, M., Bizer, C., & Paulheim, H. (2014). Adoption of the linked data best practices in different topical domains. In: International Semantic Web Conference. Springer.

[19] Sikos, L. F. (2023). Cybersecurity knowledge graphs. Knowl Inf Syst, 65, 3511–3531. https://doi.org/10.1007/s10115-023-01860-3.

[20] Szekely, P., Knoblock, C. A., Slepicka, J., Philpot, A., Singh, A., Yin, C., Kapoor, D., Natarajan, P., Marcu, D., Knight, K., et al. (2015). Building and using a knowledge graph to combat human trafficking. In: International Semantic Web Conference. pp. 205–221. Springer.

[21] Wang, S. (2017). Knowledge Graph Creation from Structure Knowledge; The Pennsylvania State University: State College, PA, USA. Shen, Y., Chen, Z., Cheng, G., & Qu, Y. (2021). CKGG: A Chinese knowledge graph for high-school geography education and beyond. In Proceedings of the Semantic Web—ISWC 2021: 20th International

Semantic Web Conference, ISWC 2021, Virtual Event 24–28 October 2021; Proceedings 20; Springer: Berlin/Heidelberg, Germany, pp. 429–445.

[22] Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014). Knowledge graph embedding by translating on hyper planes, Proceedings of the Twenty-Eighth AAAI Conference on AI, https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/viewFile/8531/8546 [accessed: 17.02.2019].

[23] Wang, S., Zhang, X., Ye, P., Du, M., Lu, Y., & Xue, H. (2019). Geographic Knowledge Graph (GeoKG): A formalized geographic knowledge representation. ISPRS Int J Geo-Inform, 8, 184.

[24] Qu, K.; Li, K.C.; Wong, B.T.M.; Wu, M.M.F.; Liu, M. A Survey of Knowledge Graph Approaches and Applications in Education. *Electronics* **2024**, *13*, 2537. https://doi.org/10.3390/electronics13132537.

[25] T. Al-Moslmi, M. Gallofré Ocaña, A. L. Opdahl and C. Veres, "Named Entity Extraction for Knowledge Graphs: A Literature Overview," in IEEE Access, vol. 8, pp. 32862-32881, 2020, doi: 10.1109/ACCESS.2020.2973928.

[26] Phillip Schneider, Tim Schopf, Juraj Vladika, Mikhail Galkin, Elena Simperl, Florian Matthes, A Decade of Knowledge Graphs in Natural Language Processing: A Survey, https://doi.org/10.48550/arXiv.2210.00105.

[27] Sanjay Budhdeo, Joe Zhang, Yusuf Abdulle, Paul M Agapow, Douglas GJ McKechnie, Matt Archer, Viraj Shah, Eugenia Forte, Ayush Noori, Marinka Zitnik, Hutan Ashrafian, Nikhil Sh arma, Scoping review of knowledge graph applications in biomedical and healthcare sciences, medRxiv 2023.12.13.23299844; doi: https://doi.org/10.1101/2023.12.13.23299844.

[28] Carl Yang, Hejie Cui, Jiaying Lu, Shiyu Wang, Ran Xu, Wenjing Ma, Yue Yu, Shaojun Yu, Xuan Kan, Chen Ling, Tianfan Fu, Liang Zhao, Joyce Ho, Fei Wang, A Review on Knowledge Graphs for Healthcare: Resources, Applications, and Promises, https://doi.org/10.48550/arXiv.2306.04802.

[29] Bingcong Xue and Lei Zou, Knowledge Graph Quality Management: A Comprehensive Survey, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 35, NO. 5, MAY 2023, https://mod.wict.pku.edu.cn/docs/20240422170853958320.pdf.

[30] Liu, K.; Wang, F.; Ding, Z.; Liang, S.; Yu, Z.; Zhou, Y. Recent Progress of Using Knowledge Graph for Cybersecurity. *Electronics* **2022**, *11*, 2287. https://doi.org/10.3390/electronics11152287.

[31] Singh, A., Verma, R., & Desai, K. (2023). Enhancing accuracy in violence detection through multi-model integration in surveillance systems. International Journal Of Recent Advances in Engineering & Technology, 11(3), 134-145.

[32] Taylor, S., & Verma, A. (2023). An Optimized Framework for Sentiment Analysis in Twitter Data using Hybrid Machine Learning Models. International Journal on Advanced Electrical and Computer Engineering, 10(1), 111-121.

[33] Smith, J., & Kumar, A. (2023). Enhancing Sentiment Analysis using Hybrid Neural Networks with Optimized Feature Selection. International Journal of Electrical, Electronics and Computer Systems, 11(2), 45-55.

[34] Bansal, M., Reddy, A., & Iyer, N. (2022). Energy-efficient framework for real-time video processing in smart cities. International Journal on Advanced Electrical and Computer Engineering, 9(4), 78-89.