# Applications of Knowledge Graphs in Biomedicine and Finance: A Survey

www.surveyx.cn

## Abstract

Knowledge graphs (KGs) have emerged as pivotal tools in transforming unstructured data into structured, machine-readable formats, facilitating enhanced data integration and analysis across various domains, notably biomedicine and finance. This survey examines the applications of KGs, emphasizing their role in improving decision-making and insights through the integration of semantic web technologies and graph databases. In biomedicine, KGs support drug discovery, disease modeling, and personalized medicine by integrating diverse datasets and enabling advanced analytical techniques. In finance, they enhance fraud detection, investment management, and risk assessment by providing a structured framework for analyzing complex relationships and integrating diverse data sources. The integration of semantic web technologies with graph databases enhances data interoperability and retrieval, while ontology-based data integration ensures semantic consistency. Despite these advancements, challenges remain in scalability, data integration, and query optimization. Future research should focus on enhancing semantic web technologies, developing flexible ontology frameworks, and integrating emerging machine learning techniques to unlock the full potential of KGs. These efforts are crucial for advancing the capabilities of KGs, supporting more robust data integration, and facilitating informed decision-making across complex domains.



## 1 Introduction

### 1.1 Concept of Knowledge Graphs

Knowledge graphs (KGs) serve as a foundational framework for converting unstructured data into structured, machine-readable formats, thereby enabling the development of intelligent applications [1]. They effectively represent entities and their interrelations, transforming raw data into actionable insights [2]. This transformation is crucial in the Semantic Web, where understanding hierarchical and aspect relationships enhances data semantic organization [3].

The integration of knowledge graphs with multimodal learning broadens their applicability, addressing knowledge gaps and providing fresh perspectives in knowledge graph research [4]. Recent surveys underscore the necessity of evolving and preserving knowledge graphs to ensure accurate and current knowledge representations in a rapidly changing environment [5].

In scientific literature, knowledge graphs facilitate structured representations of entities and their relationships, greatly improving data integration and analysis [6]. Advancements in reasoning over knowledge graphs, such as the use of dependent type theory to replicate RDF and SPARQL functionalities, further reinforce their role in data representation [7].

In cultural heritage, particularly in iconography and iconology, knowledge graphs highlight significant gaps in representing iconographic and iconological statements, often lacking granularity and under-representing cultural symbolism [8]. The semantic linked data model is essential for transforming data into information and knowledge, emphasizing the critical role of KGs in this process [9].

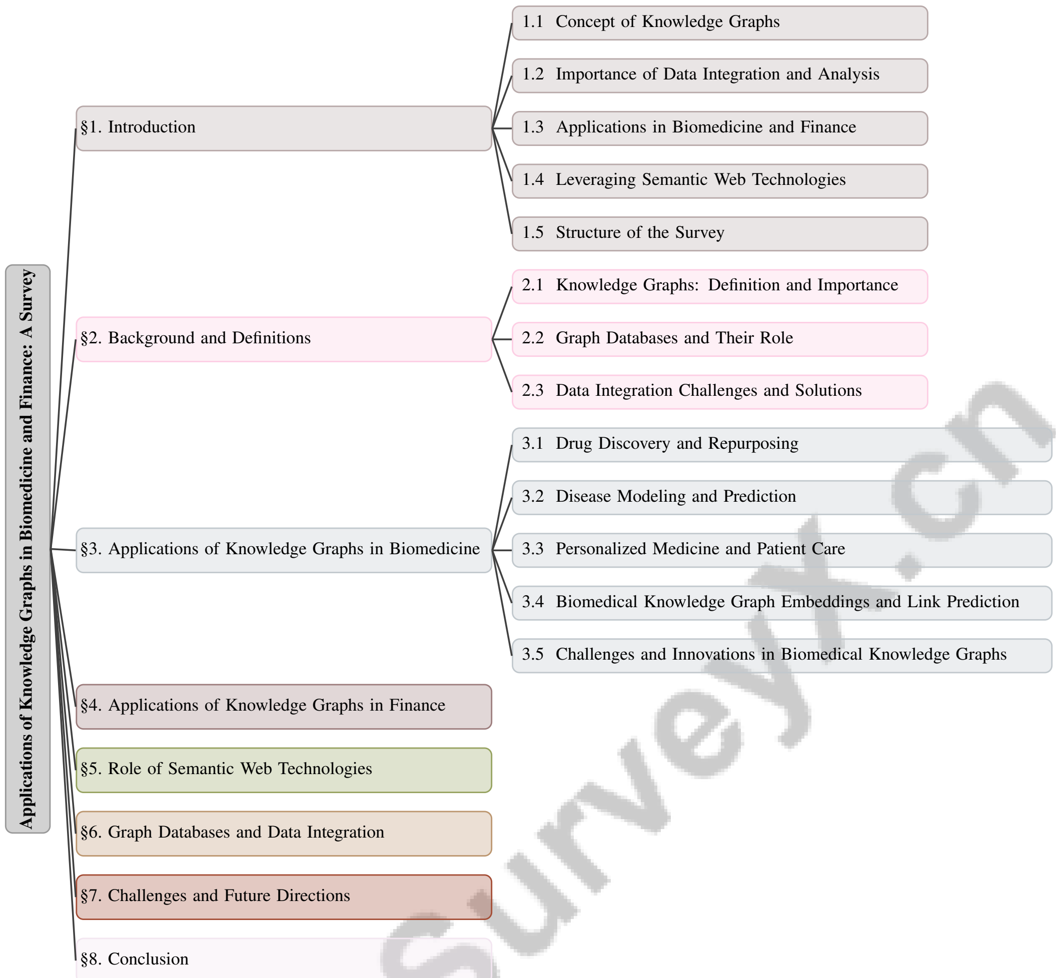


Figure 1: chapter structure

Within the broader context of artificial intelligence and big data, knowledge graphs present both opportunities and challenges as they integrate into complex systems [10]. This integration necessitates robust frameworks that support knowledge classification and organization, facilitating effective knowledge interaction [11].

### 1.2 Importance of Data Integration and Analysis

Data integration and analysis via knowledge graphs are vital for navigating the complexities of heterogeneous, large-scale datasets, particularly in biomedicine and finance. Knowledge graphs provide a robust framework to tackle semantic heterogeneity, a significant barrier to effective querying and integration across diverse sources [11]. This capability is crucial in environments where dynamic type class discovery enhances data integration adaptability [12].

In biomedicine, the integration of machine learning algorithms with knowledge graphs significantly impacts clinical outcomes by enabling nuanced data analysis and integration [12]. However, existing knowledge graph representations often fall short in providing provenance and explainability for the relationships they encode [7]. This limitation highlights the need for enhanced semantic web technologies and knowledge organization systems (KOSs) to manage and link knowledge amidst overwhelming information [11].

Graph databases are pivotal in data integration and analysis, facilitating the detection of collaboration communities among researchers, which reveals important collaboration patterns and enhances understanding of complex data structures [13]. Nevertheless, individuals without expertise in the Semantic

Web frequently struggle to access and comprehend Linked Data, resulting in information overload and hindering effective resource utilization [9].

The integration and analysis of data through knowledge graphs not only improve knowledge organization and linking but also enhance AI application performance by providing better data representation. However, challenges remain, particularly in converting natural language queries into graph query language formats compatible with domain-specific graph databases in sectors like finance and medicine [7]. Addressing these challenges is essential for maintaining knowledge graph accuracy and relevance amidst continuous real-world changes, ensuring consistency and providing provenance for updates [7].

### 1.3 Applications in Biomedicine and Finance

Knowledge graphs (KGs) are indispensable in biomedicine and finance, offering a structured framework for representing complex relationships and integrating diverse datasets. In biomedicine, knowledge graphs like the Clinical Knowledge Graph (CKG) integrate multiple databases and scientific literature with omics data, advancing discovery science and clinical practice [14]. PrimeKG exemplifies this integration by combining various biomedical resources to elucidate diseases and their interconnections with drugs, phenotypes, and biological processes [15]. Such integration is crucial for enhancing biomedical informatics and improving the performance of pre-trained language models in natural language processing tasks [16].

Knowledge graphs significantly enhance disease prediction, such as diabetes, by integrating heterogeneous gene expression datasets and domain-specific knowledge [17]. The incorporation of social media data into biomedical research, particularly for extracting focused digital cohorts in conditions like epilepsy, further demonstrates the versatility of knowledge graphs [18]. New approaches that integrate various data sources have addressed prior limitations in data exploration on dense knowledge graphs, enhancing knowledge mining in biomedical research [19].

In finance, knowledge graphs are critical for integrating causal analysis and property graphs. Causal directed acyclic graphs (cDAGs) are essential for understanding complex financial systems [20]. Frameworks like Knowledge Graph Induction (KGI) enhance recommendation and trend analysis capabilities by combining structured and unstructured data [21]. These applications highlight the potential of knowledge graphs to transform financial data into actionable insights, aiding decision-making processes.

The integration of semantic web technologies, ontologies, and linked data principles is essential in both biomedicine and finance, enabling the representation and querying of complex datasets [22]. These technologies facilitate risk estimation and support various healthcare applications, while in finance, they enable dynamic resource allocation and trend analysis, enhancing financial systems' adaptability and responsiveness. Graph database technologies, as surveyed from an industry perspective, provide the necessary infrastructure to support these applications, emphasizing the importance of robust architectures in managing and querying large-scale graph data [23].

The application of knowledge graphs in academia is exemplified through the detection of collaboration communities, enhancing understanding of relationships between researchers and their works [13]. Furthermore, exploring methodologies, challenges, and opportunities associated with Patient-Centric Knowledge Graphs (PCKGs) underscores their role in integrating disparate healthcare data and enhancing patient care through a unified health perspective [24]. The use of knowledge graphs in biomedicine and finance illustrates their capacity to integrate diverse data sources, enhance representation, and provide deeper insights into complex systems, ultimately improving decision-making and strategic planning across these critical domains.

### 1.4 Leveraging Semantic Web Technologies

Semantic Web technologies (SWT) are crucial for enhancing the functionality and accessibility of knowledge graphs by providing a framework for data interoperability and integration across diverse domains. The toolkit SemTK exemplifies this by simplifying interaction with semantic data, making it more accessible to domain experts and application developers [25]. This accessibility is vital for integrating heterogeneous data sources into coherent knowledge graphs.





AI-generated, for reference only.

The Semantic Web facilitates data interoperability, a critical requirement in complex systems like social networking applications, by leveraging ontologies and linked data principles [26]. This capability is further explored in frameworks that categorize existing research and applications, illustrating how SWT supports data interoperability and integration across various fields [27]. However, the development and evaluation of Semantic Web technologies often lack standardization, influenced by individual researchers' experiences and preferences [28].

In digital engineering, SWT provides a robust foundation for addressing integration and interoperability needs through frameworks like the Digital Engineering Framework for Integration and Interoperability (DEFII) [29]. These technologies enable efficient querying and retrieval of data from knowledge graphs, as evidenced by SparqLog, which evaluates SPARQL 1.1 queries via Datalog [30]. Such advancements highlight the potential of SWT to enhance knowledge graph querying capabilities, making them more efficient and user-friendly.

Moreover, the accessibility of semantic data to non-experts is significantly improved through web applications that allow users to query knowledge bases like DBpedia. These applications provide graphical visualizations of query results, democratizing access to complex data structures [9]. This democratization is essential for expanding the utility of knowledge graphs beyond traditional expert users, facilitating broader engagement with semantic technologies.

Semantic Web technologies are integral to the advancement and utility of knowledge graphs, providing essential tools and frameworks for data integration, interoperability, and user accessibility. As these technologies evolve, their significance in improving the functionality and accessibility of knowledge graphs across diverse sectors—such as scientific research, corporate innovation, and artificial intelligence—will likely become increasingly critical, enabling more efficient data management, enhanced analytical capabilities, and better-informed decision-making processes [10, 6, 21, 1].

### 1.5 Structure of the Survey

This survey is meticulously organized to provide a comprehensive examination of the applications of knowledge graphs in biomedicine and finance, emphasizing the integration of semantic web technologies and graph databases. The paper begins with an introduction that sets the stage by explaining the fundamental concepts of knowledge graphs and their pivotal role in data integration and analysis. The background and definitions section then provides essential definitions and explanations of core concepts such as knowledge graphs, semantic web, graph databases, and data integration, establishing a foundational understanding for readers.

The survey is divided into two main application sections: biomedicine and finance. The biomedicine section explores the use of knowledge graphs in drug discovery, disease modeling, personalized medicine, and related areas, supported by case studies and examples from the literature. The financial applications section examines how knowledge graphs are utilized in risk assessment, fraud detection, investment strategies, and more, highlighting their transformative impact on the financial sector.

Subsequent sections delve into the role of semantic web technologies, discussing advanced querying techniques, semantic representation, and integration with graph databases. The survey also analyzes the role of graph databases in managing knowledge graphs, addressing data integration challenges, and proposing solutions for interoperability and scalability.

The final section identifies current challenges and future directions in applying knowledge graphs, offering insights into potential research avenues and technological advancements. The paper concludes by summarizing key findings and reinforcing the significance of knowledge graphs in enhancing data integration and decision-making in biomedicine and finance.The following sections are organized as shown in Figure 1.

## 2 Background and Definitions

### 2.1 Knowledge Graphs: Definition and Importance

Knowledge graphs (KGs) are sophisticated data structures that represent entities and their interrelations through nodes and edges, forming a graph-based model that captures the semantics of complex datasets [31]. This structured format is crucial for organizing, querying, and integrating data across various domains, such as biomedicine and finance, addressing the challenge of leveraging



AI-generated, for reference only.

the vast information available on the Web [32]. By transforming unstructured data into structured, machine-readable formats, KGs facilitate the development of intelligent applications and clarify data, information, and knowledge relationships [7].

In biomedicine, Patient-Centric Knowledge Graphs (PCKGs) are instrumental in integrating diverse healthcare data, enabling personalized patient care [24]. They propel research in genomics and personalized medicine by integrating heterogeneous gene expression data, often constrained by sample sizes and variability [12]. The evolution and maintenance of KGs are vital for ensuring accurate knowledge representations in a rapidly changing environment [32].

In finance, KGs model complex relationships and integrate diverse datasets, enhancing decision-making processes. They provide a structured framework for understanding financial systems and their interdependencies, aiding in risk assessment, fraud detection, and investment strategy development [31]. The ability to generate realistic graph structures reflecting real-world properties is crucial for testing algorithms and applications [31].

Semantic Web technologies augment KGs by integrating statistical relational learning, enabling reasoning over extensive data across fields such as food, energy, and water systems [31]. However, semantic interoperability challenges persist between web applications and services, hindering effective communication and data exchange [9]. Tools like IICONGRAPH address issues of insufficient granularity and representation of cultural symbolism in existing KGs, particularly regarding artistic interpretations [8].

KGs are pivotal in addressing the information explosion, enhancing user experiences in online applications, and improving data-driven decision-making across sectors. As KGs evolve, their role in integrating, analyzing, and extracting insights from complex datasets is expected to grow, underscoring their significance in contemporary and future data management paradigms [32].

### 2.2 Graph Databases and Their Role

Graph databases are essential for managing and querying knowledge graphs, optimized for the complex and interconnected nature of graph data. Unlike traditional relational databases, which struggle with intricate relationships and large data volumes, graph databases like Neo4j, Amazon Neptune, and ArangoDB offer advantages by employing graph models instead of relational tables [33]. These databases excel at representing and querying complex relationships and metadata, surpassing RDF's limitations in capturing such intricacies [34].

The architecture of graph databases is divided into native and hybrid solutions, each offering distinct performance and scalability benefits [23]. Native graph databases are tailored for graph data structures, enhancing the efficiency of graph operations and queries, while hybrid solutions integrate traditional database features to address broader data management needs. Besta et al. [35] provide a comprehensive analysis of various graph database systems, including triple stores, tuple stores, native graph databases, and object-oriented systems, emphasizing their diverse functionalities and applications.

Graph databases are adept at managing highly interconnected, evolving data, crucial for applications like community detection in collaboration networks [13]. Their flexibility is further demonstrated by models like HG(2), which combine hypergraphs, graphs, and sets of connectors to represent complex problem spaces [36].

Despite their advantages, graph databases face challenges related to schema flexibility and data integration. The lack of a unified approach in current database languages complicates knowledge graph management, especially when integrating diverse data sources and ensuring schema flexibility [37]. Moreover, existing methods for addressing inconsistencies in relational databases are inadequate for the unique complexities of graph databases, necessitating new strategies for defining and computing preferred repairs [38].

Incorporating semantic web ontologies as foundational elements in graph databases enhances their role in knowledge management [39]. These ontologies facilitate the semantic representation and querying of graph data, contributing to the robustness and utility of graph databases in managing knowledge graphs. Furthermore, Bonifati et al. [40] introduce a taxonomy categorizing graph generators based on application domains and key features, aiding researchers and practitioners in selecting appropriate tools.

## 2.3 Data Integration Challenges and Solutions

Data integration within knowledge graphs (KGs) involves multifaceted challenges due to the complexity and heterogeneity of data sources. Constructing high-quality KGs is a primary concern, particularly in drug discovery, where data source reliability is crucial for accurately representing intricate relationships among genes, diseases, and drugs [41]. Comparative analyses of existing public drug discovery KGs highlight the need for data sources that capture complex relationships while maintaining high data quality [41].

In large-scale networks, inefficient resource allocation complicates data integration, resulting in suboptimal performance and increased operational costs [42]. The inability of current methods to adaptively manage resource distribution in real-time exacerbates these issues, limiting effectiveness in dynamic environments [42]. Furthermore, securely integrating diverse data sources while ensuring data quality and user access control remains a significant challenge, as the lack of secure sharing methods can lead to unauthorized access and data breaches, jeopardizing data integrity [43].

The absence of standardized practices for semantic annotation further impedes semantic alignment across domains [44]. This challenge is compounded by the manual, inefficient ontology enrichment process, which lacks scalability and complicates the integration of the growing volume of web data [31]. Integrating large, heterogeneous datasets into KGs also faces obstacles related to the volume, variety, and veracity of big data [45]. Additionally, the potential for data vandalism and misrepresentation necessitates robust mechanisms for maintaining and updating KGs effectively [5].

Ontology-based information integration emerges as a promising solution to these challenges, enhancing the identification of correspondences between entities and facilitating more effective data integration. This approach streamlines the construction and querying of KGs, enabling seamless data handling across diverse domains. Innovative solutions are required to compute preferred repairs for graph databases that do not meet specific integrity constraints, as defined by Reg-GXPath expressions [38]. Aligning large language models with graph databases also necessitates developing domain-specific natural language to graph query language data pairs, enhancing interactions between these models and graph databases [46].

In recent years, the application of knowledge graphs in biomedicine has garnered significant attention due to their potential to transform various aspects of healthcare and research. This transformation is particularly evident in several key areas, including drug discovery, disease modeling, and personalized medicine. Figure 2 illustrates the hierarchical structure of these applications, effectively highlighting not only the primary categories but also their respective subcategories and details. The figure emphasizes the integration of complex datasets and advanced learning models, as well as patient-centric approaches that are crucial for addressing the unique challenges faced in biomedical contexts. Furthermore, it showcases innovative solutions aimed at enhancing scalability and performance optimization within biomedical knowledge graphs. This comprehensive overview underscores the dynamic interplay between the advancements in knowledge graph technology and their practical applications in improving healthcare outcomes.

# 3 Applications of Knowledge Graphs in Biomedicine

## 3.1 Drug Discovery and Repurposing

Knowledge graphs (KGs) are pivotal in drug discovery and repurposing by integrating complex biomedical datasets, allowing for the identification of novel therapeutic opportunities through link prediction tasks [47]. Hetionet exemplifies the integration of diverse public data sources, enhancing drug repurposing strategies [47]. KGs unify heterogeneous data sources, crucial for managing biological data uncertainty and optimizing resource allocation, inspired by decentralized decision-making in biological systems [42]. Algorithms leveraging decentralized control mechanisms dynamically optimize resource allocation, enhancing drug discovery workflows [42].

Advanced knowledge graph embedding (KGE) models predict missing links within biomedical KGs, identifying potential drug-disease associations. However, topological imbalance can skew predictions, necessitating balanced datasets for improved accuracy [47]. Integrating fine-tuned scientific language models with traditional KG embedding models shows promise in discovering novel drug-disease associations, supporting drug repurposing efforts.

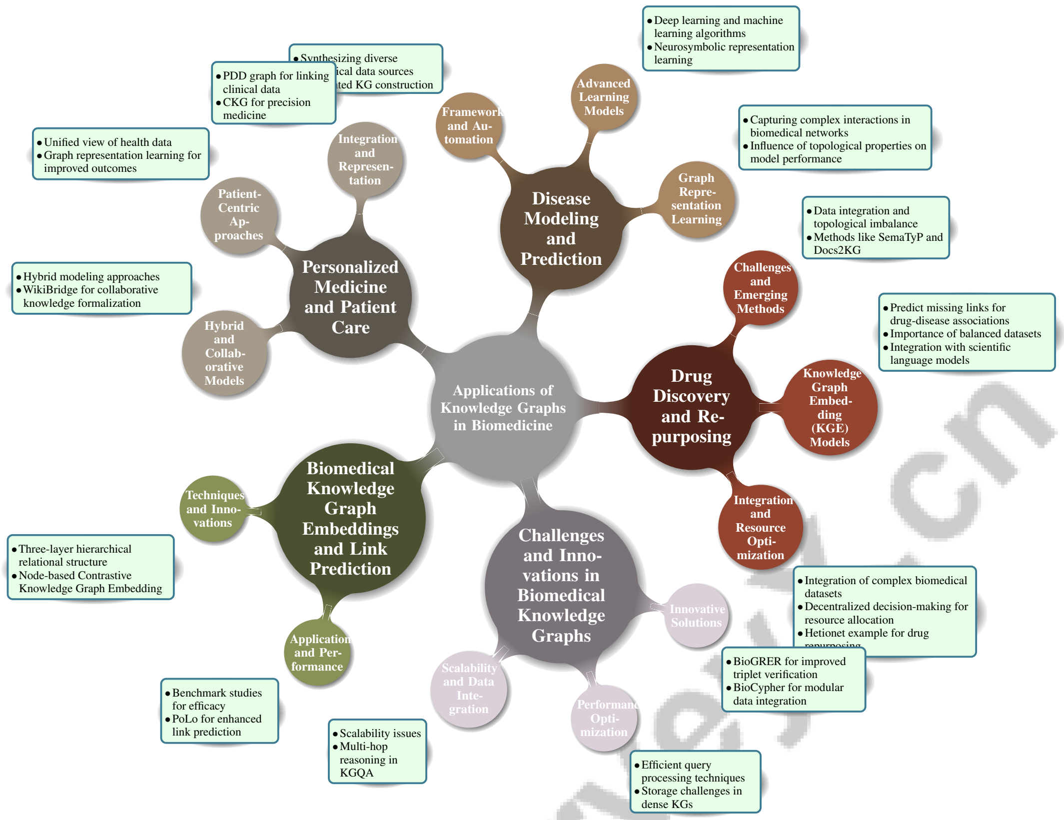


Figure 2: This figure illustrates the hierarchical structure of applications of knowledge graphs in biomedicine, highlighting key areas such as drug discovery, disease modeling, personalized medicine, and the challenges and innovations within biomedical knowledge graphs. Each primary category is further divided into subcategories and details, emphasizing the integration of complex datasets, advanced learning models, patient-centric approaches, and innovative solutions for scalability and performance optimization.

KGs provide robust infrastructure for drug discovery and repurposing, enabling the synthesis and analysis of complex biomedical data. Addressing challenges such as data integration, topological imbalance, and resource allocation will be crucial for maximizing KGs' potential in therapeutic development and personalized medicine strategies. Emerging methods like SemaTyP and Docs2KG enhance drug discovery and data management across heterogeneous biomedical literature [6, 48, 49, 47, 17].

### 3.2 Disease Modeling and Prediction

KGs are essential in disease modeling and prediction, capturing complex interactions within biomedical networks, including disease co-morbidities [50]. As illustrated in Figure 3, the hierarchical structure of disease modeling and prediction highlights key areas such as graph representation, machine learning techniques, and the challenges associated with knowledge graph construction. This figure emphasizes the critical role of biomedical networks and topological properties in enhancing predictive models.

Graph representation learning techniques integrated with KGs open new avenues for exploring these interactions in biomedical research [50]. Predicting missing links within biomedical KGs is challenging, as topological properties can influence predictive model performance [51]. Benchmarks evaluating topological features' impact on prediction accuracy enable researchers to fine-tune models [51].

Advanced deep learning models and traditional machine learning algorithms utilize semantic and structural information embedded within KGs to infer new relationships and predict disease outcomes



AI-generated, for reference only.

more accurately [52]. Neurosymbolic representation learning infers novel associations within biological KGs, crucial for advancing disease prediction and understanding health condition mechanisms [53].

KGs in disease modeling provide a sophisticated framework for synthesizing diverse biomedical data sources, enhancing the comprehension of intricate biological systems and significantly improving predictive model precision. Automating KG construction allows efficient management of vast unstructured data, improving analysis and forecasting capabilities in biomedicine [6, 1]. Addressing graph topology challenges and integrating state-of-the-art machine learning techniques will advance KGs' capabilities in biomedical research.

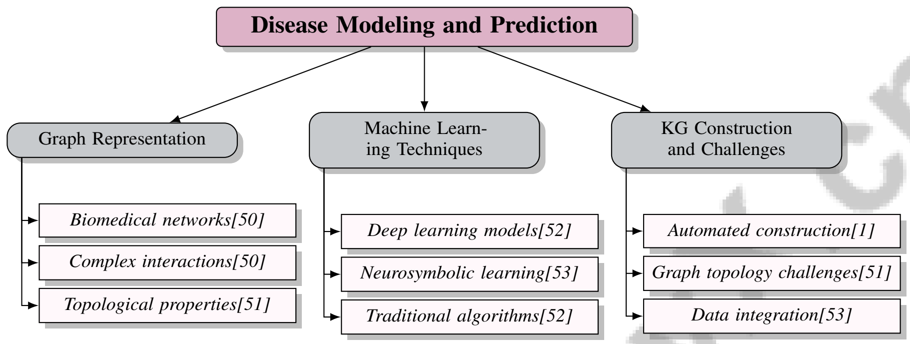


Figure 3: This figure illustrates the hierarchical structure of disease modeling and prediction, highlighting key areas such as graph representation, machine learning techniques, and knowledge graph construction challenges. It emphasizes the role of biomedical networks, complex interactions, and topological properties in enhancing predictive models.

### 3.3 Personalized Medicine and Patient Care

KGs are increasingly vital in personalized medicine and patient care, providing a comprehensive framework for integrating and analyzing diverse biomedical data. The PDD graph has significantly advanced linking clinical data with biomedical knowledge, supporting personalized medicine research by enhancing complex interaction representation [54]. This integration is crucial for developing personalized treatment strategies, allowing for seamless combinations of heterogeneous data types and facilitating better predictive analytics [55].

The Clinical Knowledge Graph (CKG) exemplifies KGs' potential in precision medicine by enabling rapid, automated proteomics data analysis, enhancing clinical decision-making and contributing to tailoring treatments to individual patients [14]. Patient-Centric Knowledge Graphs (PCKGs) offer a unified view of health data, improving disease prediction and facilitating personalized treatment strategies, ultimately enhancing patient care [24].

Graph representation learning enhances modeling of complex biomedical interactions, leading to improved outcomes in precision medicine [50]. Neurosymbolic AI methods integrated with KGs enhance interpretability and integration of rich biomedical knowledge, addressing data sparsity challenges and paving the way for improved biomedical research outcomes [56]. This approach facilitates developing more accurate and personalized diagnostic and treatment recommendations [12].

Hybrid modeling approaches combining ensemble learning with deep learning techniques improve predictive models' accuracy and generalization in personalized medicine [57]. These methods synthesize diverse data sources, supporting personalized healthcare solutions that are effective and adaptable to individual patient needs.

WikiBridge, a semantic wiki designed for collaborative knowledge formalization in life sciences, exemplifies KGs' role in enhancing patient care and personalized medicine through improved data management [58]. This collaborative approach facilitates integrating diverse knowledge sources, leading to more informed and personalized patient care strategies.

| Benchmark | Size | Domain | Task Format | Metric |
|---|---|---|---|---|
| BioKG[59] | 2,067,998 | Biomedical | Link Prediction | HITS@10, MRR |
| ESG-Graph[60] | 2,000 | Investment Management | Equity Filtering | Response Time, CPU Usage |
| RWB[52] | 100,000 | Predictive Modeling | Classification | Accuracy, F1-score |
| LiveSchema[61] | 1,000 | Knowledge Graphs | Entity Type Recognition | F1-score, Accuracy |
| Know2BIO[62] | 6,200,000 | Biomedical Knowledge Graphs | Link Prediction | Hits@k, MRR |
| DiDiC[63] | 730,027 | Geographic Information System | Graph Traversal | Inter-partition traffic |
| TIGER[47] | 2,200,000 | Biomedical Knowledge Graphs | Gene-Disease Association Prediction | MRR, Hits@10 |

Table 1: The table provides an overview of various benchmarks employed in the evaluation of knowledge graph embeddings and link prediction techniques across different domains. It details the size, domain, task format, and performance metrics for each benchmark, emphasizing their relevance in biomedical knowledge graphs and other applications.

### 3.4 Biomedical Knowledge Graph Embeddings and Link Prediction

Knowledge graph embeddings (KGE) and link prediction techniques enhance the exploration of complex biomedical data. Integrating a three-layer hierarchical relational structure (HRS) into KGE models allows richer embeddings that effectively capture relationships between semantically similar relations [64]. Node-based Contrastive Knowledge Graph Embedding (NC-KGE) advances relationship prediction by constructing appropriate contrastive pairs, improving the model's ability to discern subtle differences and similarities between entities [65]. RDF-star2Vec represents a significant advancement in KGE models by addressing RDF-star graphs, employing graph walk techniques to facilitate probabilistic transitions between query triples and their compositional entities, enhancing complex biomedical data representation [66].

Benchmark studies demonstrate KGE models' efficacy, such as ComplEx, in predicting links within biomedical KGs, showing promise for applications in drug discovery and polypharmacy [59]. PoLo, combining reinforcement learning with logical rules, enhances link prediction in biomedical KGs, leveraging machine learning and logical reasoning strengths to improve prediction accuracy and interpretability [67]. Table 1 presents a comprehensive summary of benchmarks used to assess the effectiveness of knowledge graph embeddings and link prediction methods, highlighting their application in biomedical research and other fields.

Integrating embeddings and link prediction techniques within biomedical KGs plays a crucial role in advancing biomedical research by offering sophisticated tools for modeling intricate relationships among entities and accurately predicting novel interactions. Knowledge graph embeddings significantly enhance link prediction performance, facilitating applications such as drug repurposing and drug-target identification. These embeddings enable structured information extraction from unstructured clinical data, supporting complex relationships and interactions in the biomedical domain, driving innovation in diagnosis and treatment methodologies [68, 51, 69, 59, 70]. As these techniques evolve, they hold the potential to significantly impact drug discovery, personalized medicine, and the broader understanding of biomedical systems.

### 3.5 Challenges and Innovations in Biomedical Knowledge Graphs

Biomedical knowledge graphs (KGs) face challenges regarding scalability, data integration, and efficiently representing complex biomedical entity interactions. Current methodologies often struggle with scalability, managing large, heterogeneous datasets, limiting KGs' applicability across diverse biomedical contexts [50]. Multi-hop reasoning in biomedical knowledge graph question answering (KGQA) requires comprehensive datasets capable of addressing biomedical reasoning intricacies [71].

Performance optimization is critical, as certain queries experience prolonged execution times, necessitating efficient query processing techniques [72]. The dense nature of biomedical KGs often leads to storage challenges, where the inability to store the complete graph within the database can result in performance bottlenecks [19]. Frequent updates required for maintaining large volumes of profiles in biomedical KGs further complicate their management and processing [73].

Innovations address these challenges through advanced methods for knowledge graph completion and link prediction. BioGRER combines the strengths of knowledge graph embedding and logic rules, leading to improved triplet verification performance compared to existing methods [74]. Integrating neurosymbolic AI reasoning with KGs enhances interpretability and inference capabilities, providing a robust framework for understanding complex biomedical interactions [53].

Tools like BioCypher, with their modular architecture, facilitate integrating multiple data sources, reducing barriers to knowledge graph creation and enhancing usability for non-specialists [75]. This modularity is crucial for democratizing knowledge representation and enabling broader participation in developing and using biomedical KGs. PoLo improves link prediction tasks by providing interpretable reasoning paths, essential for understanding underlying biological relationships [67].

The backbone-based approach in selecting focused digital cohorts allows for the inclusion of low-engagement users who provide valuable content, enhancing data integration within biomedical KGs [18]. However, challenges remain in managing entities with non-informative descriptions or insufficient textual context, which can hinder scientific language models' performance in biomedical applications [76].

## 4 Applications of Knowledge Graphs in Finance

Knowledge graphs (KGs) are increasingly vital in finance, addressing complex challenges across various domains, such as fraud detection and causal analysis, investment management, and risk assessment. By structuring relationships within financial systems, KGs enhance transaction integrity and security.

### 4.1 Fraud Detection and Causal Analysis

KGs are instrumental in fraud detection and causal analysis, offering a structured framework to represent intricate relationships. Integrating causal directed acyclic graphs (cDAGs) with KGs enables financial institutions to discern complex interactions and dependencies, crucial for identifying fraudulent patterns and understanding causal mechanisms [20]. Graph databases like Neo4j enhance KGs' capabilities by enabling efficient querying and data processing, outperforming traditional databases such as PostgreSQL and ElasticSearch, especially in scenarios like ESG-integrated equity filtering [60]. KGs facilitate causal analysis by integrating diverse data sources, including unstructured documents, leading to a coherent representation of financial transactions. This integration supports various downstream tasks, uncovering hidden patterns contributing to fraudulent behavior [77, 6, 48, 64, 78].

As illustrated in Figure 4, the hierarchical structure of key concepts in fraud detection and causal analysis highlights the pivotal role of knowledge graphs, the efficiency of graph databases, and advanced techniques in data integration and causal modeling. Leveraging advanced graph database technologies and causal modeling techniques, KGs transform unstructured data into machine-readable formats, enhancing process accuracy and efficiency in fraud prevention [77, 6, 10, 79, 1].

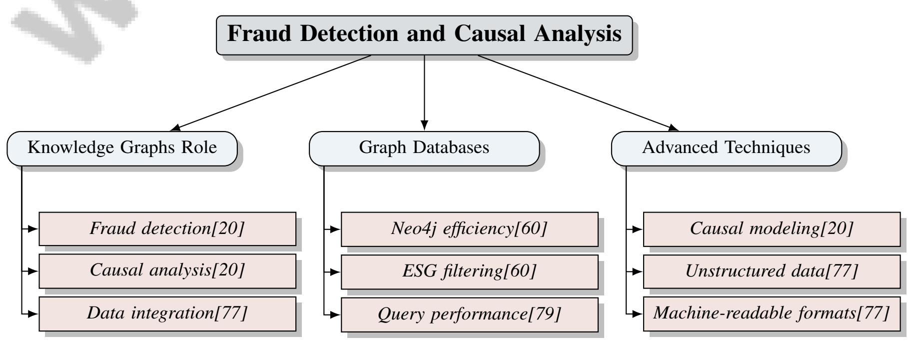


Figure 4: This figure illustrates the hierarchical structure of key concepts in fraud detection and causal analysis, focusing on the role of knowledge graphs, the efficiency of graph databases, and advanced techniques in data integration and causal modeling.

### 4.2 Investment Management and ESG Integration

In investment management and ESG integration, KGs provide a sophisticated framework for analyzing complex financial data. Graph databases, such as Neo4j, underscore their computational efficiency in managing interconnected data compared to traditional SQL and No-SQL databases [60]. KGs facilitate the aggregation and analysis of diverse data sources, enabling comprehensive assessments of the sustainability and ethical impact of investments. Their structured design allows for intricate representations of relationships between financial entities and ESG factors, enhancing equity filtering processes and supporting informed decision-making [60, 48, 64]. Advanced graph database technologies significantly improve querying efficiency, crucial for dynamic resource allocation and trend analysis, with algorithmic optimization for large-scale KGs indicating potential performance improvements [80, 60]. Integrating KGs into investment management and ESG practices enhances decision-making, transparency, and accountability, supporting the transition to sustainable financial systems [60, 48, 10, 6, 1].

### 4.3 Risk Assessment and Data Unification

KGs play a crucial role in risk assessment and data unification by providing a robust framework for integrating and analyzing diverse datasets. In investment management, efficient filtering of ESG-integrated equities is vital for assessing risks and aligning strategies with sustainability goals. Benchmarking graph databases like Neo4j illustrates their superior performance in managing interconnected data, particularly in ESG integration [60]. The structured architecture of KGs facilitates the integration of heterogeneous data sources, enhancing data unification and supporting advanced querying capabilities. Frameworks like Docs2KG dynamically construct unified knowledge graphs from unstructured documents, leading to improved decision-making [77, 6, 48, 64, 78]. KGs extend their application in risk assessment beyond investment management to fraud detection and causal analysis, enhancing the understanding of interactions and dependencies [78, 6, 48, 21]. As data management and AI evolve, the integration of graph databases and KGs will significantly influence risk management and data integration, structuring vast unstructured data into machine-readable formats and facilitating advanced applications like semantic search [37, 6, 10, 79, 1].

### 4.4 Investment Strategies and Information Retrieval

KGs significantly enhance investment strategies and information retrieval by providing a structured framework for analyzing complex financial data. Their integration into investment management processes allows for the synthesis of diverse data sources, fostering informed decision-making and robust strategy development. By representing financial entities as nodes and edges, KGs facilitate the identification of crucial patterns and trends for strategic planning [60]. In information retrieval, KGs improve search precision and relevance by leveraging semantic relationships between entities, essential in financial markets for strategic decisions. Advanced graph database technologies like Neo4j enhance this process by enabling efficient querying and retrieval of relevant data, critical for dynamic resource allocation and trend analysis [60]. The application of KGs in investment strategies also encompasses the integration of ESG factors, providing a comprehensive view of portfolios and aiding investors in aligning their portfolios with sustainability goals [60]. Integrating KGs into investment strategies and information retrieval processes offers a powerful framework for enhancing decision-making and supporting the transition to sustainable financial systems. These technologies facilitate the structuring of vast unstructured data into machine-readable formats, enabling sophisticated applications, including semantic search and deep question answering, and supporting complex problem-solving across various domains [6, 1].

## 5 Role of Semantic Web Technologies

Semantic web technologies significantly enhance querying and data retrieval processes, improving interactions with complex datasets and supporting knowledge graph applications. These technologies, through the use of ontologies and advanced querying techniques, offer deeper insights into data extraction, crucial for data-intensive sectors like life sciences, where knowledge-driven querying systems are replacing traditional methods to improve research outcomes and data accessibility [81, 27, 39].

## 5.1 Advanced Querying and Data Retrieval

Advanced querying and data retrieval are essential for leveraging semantic web technologies in efficiently interacting with complex knowledge graphs. The integration of SPARQL with RDF graphs, as demonstrated by Ondex's interactive importer, facilitates query-driven dataset construction, enhancing interoperability with the Semantic Web [82]. This is crucial for extracting insights in domains like biomedicine and finance. The mutual transformation of data, information, and knowledge, critical for systems like the Semantic Web Expert System (SWES), underscores the importance of advanced querying in semantic web applications [2]. Dependently Typed Knowledge Graphs (DTKGs) further enhance reasoning and explainability by unifying RDF and SPARQL through dependent type theory [7].

The Scientific Knowledge Graph Generation Framework employs NLP and Machine Learning to extract entities and relationships from scientific texts, improving querying capabilities and data retrieval efficiency [6]. Personal Knowledge Graphs (PKGs) in collaborative search environments introduce a new paradigm for personalization, addressing existing method gaps [83]. The effectiveness of advanced querying is further shown in SPARQL and SQL queries on Oracle databases, which utilize ranking methods based on degree decoupling for RDF instances [84]. The ARAA framework's adaptive learning optimizes information integration and query efficiency by dynamically responding to network conditions and resource demands [42].

User-specific access control in querying diverse data is achieved through blockchain-based methods, providing a unified interface that enhances protection and efficiency [43]. These advancements in querying techniques within semantic web technologies unlock the full potential of knowledge graphs, facilitating seamless data integration and analysis, and supporting informed decision-making across domains.

Figure 5 illustrates the hierarchical categorization of advanced querying and data retrieval strategies in semantic web technologies. It highlights three primary areas: Semantic Web Integration, Knowledge Graphs Enhancement, and Resource and Security Optimization, each featuring key methodologies and frameworks from recent research. This visual representation complements the discussion by providing a structured overview of the various approaches and their interconnections, thereby enhancing the reader's understanding of the complexities involved in this field.

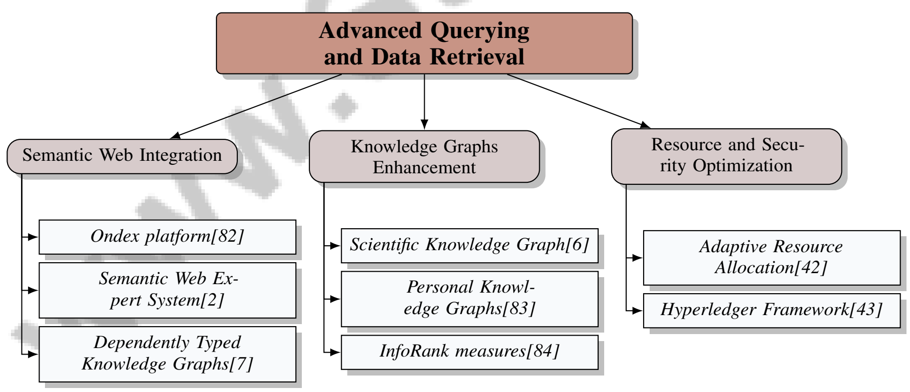


Figure 5: This figure illustrates the hierarchical categorization of advanced querying and data retrieval strategies in semantic web technologies. It highlights three primary areas: Semantic Web Integration, Knowledge Graphs Enhancement, and Resource and Security Optimization, each featuring key methodologies and frameworks from recent research.

## 5.2 Semantic Representation and Reasoning

Semantic representation and reasoning within knowledge graphs (KGs) enhance their expressiveness and utility. The integration of hierarchical relation structures in knowledge graph embeddings (KGE) models, as presented by Zhang et al., improves the embedding process for richer representations of complex relationships [64]. This is significant in biomedicine, where accurately modeling intricate interactions is crucial for effective knowledge discovery [59].

Ontology-based methods improve interpretability and accuracy in classification tasks by providing a structured framework for semantic representation, facilitating relevant feature selection [85]. Combining ontologies with rule-based systems augments knowledge representation expressiveness, enabling sophisticated reasoning capabilities in hybrid knowledge bases [86]. The Typed Graph Model (TGM) enhances semantic expressiveness by supporting complex data types and abstraction levels in data modeling [87].

In biomedicine, refining knowledge graphs through methods like BioGRER leverages semantic information from embeddings and logical constraints from rules, enhancing triplet verification accuracy [74]. The Coupled Statistical/Semantic Framework (CSSF) automates ontology enrichment using statistical and semantic analyses, enhancing semantic representation [31]. Adaptive resource management frameworks, utilizing multi-agent system architectures, provide a theoretical basis for dynamic semantic reasoning and resource allocation, crucial for maintaining knowledge graphs' relevance [88]. These advancements in semantic representation and reasoning are integral to developing and applying knowledge graphs, enabling effective data integration, analysis, and decision-making.

### 5.3 Integration with Graph Databases

Integrating semantic web technologies with graph databases enhances the functionality and adaptability of knowledge graph systems, bridging traditional databases and modern semantic standards. Frameworks like S2CTrans translate SPARQL queries into Cypher queries, preserving semantics and utilizing graph relational algebra for efficient translation [89]. This enhances interoperability between graph query languages, increasing graph databases' flexibility.

RDF2RDB-REST exemplifies integration by analyzing RDF datasets to create relational database models and generating source code for REST APIs [90]. This method transforms semantic data into formats compatible with relational databases, enhancing accessibility across platforms. Aligning large language models (LLMs) with domain-specific graph databases by generating NL-GQL pairs improves query generation accuracy, enhancing LLM and graph database interaction [46]. This alignment leverages LLM capabilities in querying complex graph data, enhancing graph databases' functionality.

Distributed architectures, like Hyperledger-based systems, serve as a view layer for data access and security management [43]. This ensures secure and efficient data management, crucial for maintaining data integrity in graph databases. Tools like IICONGRAPH demonstrate the integration of semantic web technologies, introducing a structured approach to iconographic and iconological statements, enhancing granularity and symbolic representation [8]. This is crucial for improving knowledge graphs' semantic richness and expressiveness, supporting nuanced data analysis and interpretation.

## 6 Graph Databases and Data Integration

| Category | Feature | Method |
|---|---|---|
| **Transforming RDF Data into Property Graphs** | Dynamic Conversion | OCDI[91], PRGD[38], ERGS[92], TH++[93] |
| **Schema-less and Schema-driven Approaches** | Hybrid Strategies | S2CTrans[89] |
| **Ontology-based Data Integration** | Semantic Interoperability<br>Enhanced Reasoning | DM-RDF-PG[94]<br>PoLo[71] |
| **Query Optimization and Performance** | Query Execution Strategies | MSOA[80], MWLJ[95], RAPID+[96] |
| **Integration of Diverse Data Sources** | Semantic Integration | D2KG[48], SEMO[97], SPARQL2XQuery[98] |

Table 2: This table presents a comprehensive summary of various methodologies employed in graph database transformations and data integration. It categorizes the methods based on their primary focus, such as transforming RDF data into property graphs, schema-less and schema-driven approaches, ontology-based data integration, query optimization, and integration of diverse data sources. Each method is associated with specific features and techniques, highlighting the diverse strategies utilized in enhancing graph database technologies.

The transformation of data within graph databases is critical for improving data management efficiency and effectiveness. A primary focus is the conversion of Resource Description Framework (RDF) data into property graphs, which enhances querying capabilities while preserving semantic integrity, thus promoting dynamic data integration. Table 3 provides a detailed overview of the methodologies and techniques applied in transforming RDF data into property graphs, highlighting the significance of these approaches in modern graph database technologies. The following subsection will explore the methodologies and implications of this transformation, highlighting its importance in modern graph database technologies.

### 6.1 Transforming RDF Data into Property Graphs

Transforming RDF data into property graphs represents a significant advancement in graph data management, offering greater flexibility and efficiency in querying and updating databases. This process involves direct mappings from RDF to property graph databases, maintaining semantic integrity [94]. By converting RDF triples into a property graph model, this approach facilitates intuitive data interactions and enables complex querying.

A notable advantage is the adoption of schema-less approaches, as seen in the GRAD graph database model, which supports dynamic updates without predefined schemas. This flexibility is particularly useful in fast-changing environments, such as integrating national COVID-19 data into RDF triples for spatial and temporal alignment [91]. The transformation also enhances subgraph isomorphism task efficiency, demonstrated by TurboHOM++, which converts RDF data into labeled graphs for efficient processing [93].

Beyond technical benefits, this transformation addresses data integrity challenges. Methods for computing preferred repairs based on integrity constraints ensure that transformed data remains consistent [38]. Additionally, integrating RDF datasets with graph databases, as seen in the Expressive Reasoning Graph Store (ERGS), enables SPARQL queries through a translation module, bridging RDF and property graph models [92].

This transformation offers numerous advantages, including improved flexibility, efficient querying, and enhanced data integrity. As graph database technologies advance, this transformation will be crucial for enhancing knowledge graph capabilities, facilitating efficient data integration and analysis across diverse domains. By leveraging natural language processing and machine learning techniques, knowledge graphs can extract and represent complex relationships from vast datasets, supporting applications like semantic search and personalized recommendations. The development of standardized graph query languages and automated construction methods will further reduce the costs of building high-quality knowledge graphs, enabling effective management and dissemination of scientific knowledge [77, 37, 6, 79, 1].

### 6.2 Schema-less and Schema-driven Approaches

The comparison between schema-less and schema-driven approaches in graph databases reveals their inherent flexibility and constraints, impacting graph data management and querying. Schema-less approaches, such as property graph models, allow dynamic updates and data additions without predefined schemas, advantageous in environments with frequently changing data structures [89]. Conversely, schema-driven approaches, often associated with RDF and SPARQL, require a predefined schema that dictates data structure and relationships, ensuring consistency and integrity but potentially limiting adaptability.

The primary challenge in integrating these approaches lies in the significant differences in semantic representation and processing logic between SPARQL and property graph query languages like Cypher, complicating the translation process due to distinct data handling and query execution methods [89]. The choice between schema-less and schema-driven approaches often depends on specific application requirements, balancing flexibility against the need for consistency and defined structure.

Both schema-less and schema-driven approaches present unique advantages and challenges, and the decision to utilize one over the other should be informed by the specific needs and constraints of the application domain. As graph database technologies evolve, hybrid models that integrate the strengths of both approaches may offer a promising strategy for improving the flexibility and robustness of graph data management. These hybrid systems can effectively manage complex datasets by leveraging diverse data models and query languages, enhancing data processing, analysis, and knowledge extraction efficiency across various application domains [6, 35, 79, 99, 100].

## 6.3 Ontology-based Data Integration

Ontology-based data integration employs formal ontologies to provide a unified framework for integrating heterogeneous data sources, enhancing semantic interoperability and consistency in knowledge graphs. This method utilizes ontologies to define a shared vocabulary and relationships among data entities, facilitating seamless data merging across diverse domains. Integrating ontologies with graph databases ensures that the property graph model can encompass the information capacity of the RDF model, as demonstrated by formal definitions for both schema-dependent and schema-independent mappings [94].

A key benefit of ontology-based data integration is its structured approach to reasoning over knowledge graphs, exemplified by the PoLo method, which employs policy-guided walks based on reinforcement learning, augmented by logical rules for structured reasoning [71]. By incorporating logical rules into the integration process, ontology-based methods enable more accurate inferences, improving the quality and reliability of integrated data.

Furthermore, ontology-based integration supports aligning data from disparate sources through a common semantic framework, bridging gaps between different models and formats. This alignment is crucial for ensuring data consistency and enabling complex querying and reasoning tasks across integrated datasets. Ontology-based methods are particularly beneficial in fields like biomedicine and finance, where seamlessly integrating diverse data sources is essential for extracting meaningful insights and facilitating informed decision-making. These methods leverage specialized vocabularies and semantic structures inherent in ontologies, enhancing the curation and integration of large-scale datasets, as evidenced by advancements like ConMap, which streamlines the processes of semantification, curation, and integration [101, 102].

Ontology-based data integration significantly enhances the semantic richness and interoperability of knowledge graphs, providing a robust foundation for integrating and reasoning over complex datasets. As knowledge representation and management evolve, developing sophisticated ontology-based methods will be crucial for fully leveraging knowledge graph capabilities across diverse applications, including scientific research, healthcare, and business intelligence. These knowledge graphs effectively organize vast amounts of unstructured data into structured formats of entities and relationships, enhancing functionalities like semantic search, question-answering, and advanced analytics. By employing state-of-the-art Natural Language Processing and Machine Learning techniques, researchers can automate the extraction of meaningful insights from extensive literature, improving knowledge graph construction efficiency and enabling more effective decision-making across various domains [6, 1].

## 6.4 Query Optimization and Performance

Query optimization and performance enhancement are essential for managing graph databases, particularly when handling large and complex datasets. The Multi-step Optimization Approach (MSOA) exemplifies a strategy focused on simplifying queries to improve execution efficiency, thereby reducing runtime and resource consumption [80]. This approach is particularly effective for accessing and processing large datasets, enabling more efficient data retrieval.

Research in query optimization techniques has shown significant advancements in managing complex queries and dynamic data, essential for improving graph database performance and allowing for efficient query execution and data management [103]. Optimizing complex queries is crucial in continuously evolving environments, ensuring that databases adapt to changing structures and requirements without compromising performance.

The performance of graph databases is further enhanced through native execution of GraphQL queries, assessed based on execution times and scalability across various datasets generated by LinGBM. This assessment emphasizes the importance of optimizing join and left-join operations, common in graph database queries, which can significantly impact performance if not managed effectively [95]. By focusing on these operations, graph databases can achieve better scalability and efficiency, supporting more complex applications.

RAPID+ is another notable optimization technique that enhances query processing efficiency by managing related sets of triples as first-class citizens, reducing computational overhead for faster execution [96]. By treating related datasets as integral components of the query process, RAPID+

improves overall graph database performance, particularly in applications requiring ontological query processing.

Strategies for optimizing queries and enhancing performance in graph databases are crucial for managing the increasing complexity and scale of modern data environments, especially as these databases support diverse applications such as knowledge mining in life sciences and social network analysis. Recent research highlights the importance of algorithmic optimization to address challenges in large-scale labeled property graphs, demonstrating speedups of 44 to 3839 times compared to naive querying methods. As graph databases evolve, understanding their unique characteristics and implementing advanced query optimization techniques will be essential for leveraging the full potential of graph-like data structures [79, 80, 103, 99]. As these techniques advance, they will play a crucial role in enhancing graph database capabilities, enabling more efficient data integration, retrieval, and analysis across various domains.

### 6.5 Integration of Diverse Data Sources

Integrating diverse data sources into knowledge graphs is critical for enhancing their utility across various domains, allowing for a comprehensive representation of complex datasets. A significant challenge in this integration is maintaining the integrity of original relational databases while providing a clear semantic structure that aligns with OWL, thereby facilitating better interoperability [97]. The integration process is further complicated by managing multi-level relationships within large datasets, a task that traditional relational databases struggle to handle efficiently.

To address these challenges, methodologies like Docs2KG have been developed to manage diverse unstructured data formats and dynamically integrate information, enhancing interoperability [48]. Similarly, the SPARQL2XQuery Framework allows users to work solely with SPARQL without needing to understand XML Schema complexities, simplifying the querying process [98]. This simplification is crucial for broader access to knowledge graph technologies and integrating heterogeneous data sources.

The integration of diverse data sources is effectively illustrated through the MIMIC-III clinical care database, comprising 505 patient clinical notes. This example highlights the potential of Semantic Web technologies to facilitate seamless data integration across heterogeneous healthcare sources. By leveraging ontologies and semantic integration techniques, as proposed in recent studies, it becomes feasible to enhance knowledge exchange, improve data interoperability, and support decision-making processes within healthcare systems, addressing the challenges of managing vast volumes of clinical information [104, 22, 101]. This underscores the importance of semantic technologies in enhancing data harmonization and interoperability across diverse domains.

Additionally, the proposed method for mapping relational databases to OWL maintains the integrity of original data while providing a semantic structure that enhances interoperability [97]. This method is essential for ensuring data consistency and enabling complex querying and reasoning tasks across integrated datasets.

| Feature | Transforming RDF Data into Property Graphs | Schema-less and Schema-driven Approaches | Ontology-based Data Integration |
|---|---|---|---|
| **Integration Approach** | Direct Mapping | Hybrid Models | Ontology Framework |
| **Flexibility** | Schema-less Updates | Dynamic Updates | Semantic Interoperability |
| **Performance** | Efficient Querying | Varies BY Schema | Improved Reasoning |

Table 3: This table provides a comparative analysis of three methodologies for transforming RDF data into property graphs: schema-less updates, schema-driven approaches, and ontology-based data integration. It highlights the integration approaches, flexibility, and performance characteristics of each method, offering insights into their respective advantages and limitations within graph database technologies.

## 7 Challenges and Future Directions

The evolution of knowledge graphs necessitates exploring future directions to enhance their applications. Central to this is advancing semantic web technologies to overcome current challenges and fully realize knowledge graph potential. This involves developing adaptable ontology frameworks for improved interoperability and integrating advanced technologies to enhance semantic web systems'



AI-generated, for reference only.

functionality. Ontologies serve as foundational structures for data annotation, facilitating a shift from a document-oriented to a data-oriented web enriched with semantics. Such transformation is crucial for data-intensive industries, enabling automated processes like service discovery and composition, and addressing dynamic ontology enrichment to meet evolving user needs [31, 27, 39].

### 7.1 Enhancing Semantic Web Technologies

Enhancements in semantic web technologies are vital for addressing existing challenges and maximizing knowledge graphs' potential. Future research should focus on developing adaptable ontology frameworks and improving ontology mapping capabilities to foster semantic interoperability. The Semantic Web relies on ontologies for a common vocabulary, enabling automated annotation, discovery, and service composition. Addressing the dynamic enrichment of ontologies with current information and user requirements is critical for bridging relational knowledge gaps and enhancing semantic technologies [31, 39]. This includes exploring geometric modeling methods and integrating ontology embeddings with large language models to advance neural-symbolic integration.

Refining frameworks like S2CTrans to support additional SPARQL features and exploring reverse translation from Cypher to SPARQL are essential for enhancing interoperability between query languages [89]. Secure access control mechanisms integrating heterogeneous data sources are crucial for data privacy and protection [43].

Emerging trends focus on robust frameworks for ethical AI in healthcare and improved system interoperability. Advancements in middleware solutions for tool-specific data transformations and integrating diverse analytical models into comprehensive frameworks enhance semantic web technologies' adaptability and scalability. These developments improve interoperability of independently produced information, enabling efficient data integration, visualization, and analysis across domains. Leveraging ontologies as foundational elements allows richer semantic annotations and coherent understanding of complex systems, fostering machine-readable and understandable web content [31, 105, 44, 39]. Optimizing database infrastructure and exploring alternative graph databases are critical for managing dense knowledge graphs.

Addressing challenges in graph databases, such as improving transaction support, enhancing data import/export processes, and integrating graph technologies with non-graph workloads, is vital for expanding semantic web technologies' applicability. Refining the GQL standard and investigating emerging trends in graph database technologies will facilitate seamless data integration and enhance interoperability between relational and graph databases. Efforts to address complexities of adding new metadata types to graph structures, like edge-labelled graphs, and adopting flexible models like property graphs, increasingly supported by major graph database engines, will improve data querying and representation efficiency [35, 106, 79, 99, 95].

Future research should prioritize developing advanced interfaces for natural language processing to enhance query formulation capabilities and explore innovative methods to increase knowledge-driven querying systems' efficiency. This focus is essential given the growing volume of scientific literature, complicating research data analysis and management. Utilizing state-of-the-art natural language processing and machine learning techniques, researchers can create effective knowledge graphs that accurately represent entities and relationships within scholarly publications. Investigating the comparative advantages of knowledge-driven querying systems over traditional database approaches can facilitate intuitive translations of researchers' questions into computational queries, improving data retrieval and analysis in life sciences and beyond [81, 6]. Enhancing SPARQL feature coverage and optimizing query plans specific to SPARQL are crucial for developing a unified benchmark for comprehensive testing.

Emphasizing practical implementations of semantic web technologies in healthcare is essential, focusing on developing interoperability frameworks and addressing data integration and security challenges. Improving indexing strategies, optimizing cache management, and broadening reasoning capabilities to accommodate more complex OWL profiles will enhance semantic web technologies' expressiveness and utility. These advancements will enable more meaningful machine-readable information representation, facilitating automated processes like service discovery and composition, transforming the web into a more machine-understandable framework. Leveraging ontology as a foundational element will support creating a common vocabulary and semantic descriptions that enhance data interoperability across diverse domains, including bioinformatics and knowledge management [107, 81, 39].

### 7.2 Advancements in Machine Learning and AI Integration

Integrating machine learning (ML) and artificial intelligence (AI) with knowledge graphs (KGs) is transforming complex datasets' analysis and utilization across domains. A significant focus is on developing advanced ontological frameworks to manage large-scale data and support real-time feature selection processes [85]. Exploring ontology embeddings with large language models offers promising advancements in representing complex knowledge structures and enhancing KGs' semantic richness [108]. This integration is crucial for stratified data integration, where independently managing different types of heterogeneity can significantly enhance integration outcomes [78].

Machine learning techniques, such as the Edge Confidence method, quantify similarity between predicates using lexical similarity, allowing broader inclusion of edges in the propagation graph, thus enhancing KGs' connectivity and utility [109]. Future research should focus on optimizing algorithms for computing statistics within KGs, vital for advancing applications [110]. Developing sophisticated graph generators that can be highly configurable to adapt to specific application needs is a critical exploration area [40].

ML and AI integration in managing hybrid knowledge bases is exemplified by methods like PoLo, incorporating logical rules into reinforcement learning frameworks [71]. This approach highlights potential enhancements in applying logical rules to biomedical tasks. Exploring error approximations for larger sets of additional nodes and different graph structures could influence centrality measures' stability, improving KG applications' robustness [111].

Despite these advancements, comprehensive frameworks to address ontology mapping and integration complexities remain necessary [39]. Integrating statistical techniques for acquiring missing semantic relations can enhance the ontology merging process, supporting more effective ML and AI applications in KGs [112]. Expanding Personal Knowledge Graphs (PKGs) applications across domains underscores the importance of improving collaborative features and addressing privacy concerns through legal collaborations [83].

Future work should focus on developing a user-friendly query language based on the functional model and creating a robust prototype for real-world applications [106]. Addressing current challenges and leveraging technological advancements, integrating machine learning and AI with knowledge graphs presents a promising avenue for enhancing data analysis and decision-making capabilities, unlocking new opportunities for utilizing KGs' full potential in complex data environments.

## 8 Conclusion

Knowledge graphs have emerged as pivotal tools in enhancing data integration and decision-making processes within biomedicine and finance. By utilizing RDF standards, they ensure interoperability and enable efficient data exchange, which is particularly beneficial in complex domains. In biomedicine, knowledge graphs significantly improve predictive capabilities, such as in diabetes prediction, by integrating diverse datasets and overcoming challenges like limited sample sizes. Their ability to identify semantically related entities with high precision and recall demonstrates their effectiveness in managing intricate data environments.

Despite these advancements, the application of Semantic Web technologies in healthcare presents ongoing challenges, especially in real-world implementation and validation. Addressing these challenges requires continuous research and development to refine these technologies. The successful translation of SPARQL queries to Cypher through innovative frameworks highlights potential improvements in query performance for property graph databases, suggesting promising research directions.

Future endeavors should focus on advancing semantic web technologies, optimizing graph database infrastructures, and enhancing interoperability among various data models. These efforts are essential for furthering the capabilities of knowledge graphs, ensuring their successful application across complex fields, and supporting comprehensive data integration and analytical processes.

## References

[1] Ali Hur, Naeem Janjua, and Mohiuddin Ahmed. A survey on state-of-the-art techniques for knowledge graphs construction and challenges ahead. In *2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 99–103. IEEE, 2021.

[2] Olegs Verhodubs. Mutual transformation of information and knowledge, 2016.

[3] Aida Slavic. Classification revisited: a web of knowledge, 2017.

[4] Zhuo Chen, Yichi Zhang, Yin Fang, Yuxia Geng, Lingbing Guo, Xiang Chen, Qian Li, Wen Zhang, Jiaoyan Chen, Yushan Zhu, Jiaqi Li, Xiaoze Liu, Jeff Z. Pan, Ningyu Zhang, and Huajun Chen. Knowledge graphs meet multi-modal learning: A comprehensive survey, 2024.

[5] Nacira Abbas, Kholoud Alghamdi, Mortaza Alinam, Francesca Alloatti, Glenda Amaral, Claudia d'Amato, Luigi Asprino, Martin Beno, Felix Bensmann, Russa Biswas, Ling Cai, Riley Capshaw, Valentina Anita Carriero, Irene Celino, Amine Dadoun, Stefano De Giorgis, Harm Delva, John Domingue, Michel Dumontier, Vincent Emonet, Marieke van Erp, Paola Espinoza Arias, Omaima Fallatah, Sebastián Ferrada, Marc Gallofré Ocaña, Michalis Georgiou, Genet Asefa Gesese, Frances Gillis-Webber, Francesca Giovannetti, Marìa Granados Buey, Ismail Harrando, Ivan Heibi, Vitor Horta, Laurine Huber, Federico Igne, Mohamad Yaser Jaradeh, Neha Keshan, Aneta Koleva, Bilal Koteich, Kabul Kurniawan, Mengya Liu, Chuangtao Ma, Lientje Maas, Martin Mansfield, Fabio Mariani, Eleonora Marzi, Sepideh Mesbah, Maheshkumar Mistry, Alba Catalina Morales Tirado, Anna Nguyen, Viet Bach Nguyen, Allard Oelen, Valentina Pasqual, Heiko Paulheim, Axel Polleres, Margherita Porena, Jan Portisch, Valentina Presutti, Kader Pustu-Iren, Ariam Rivas Mendez, Soheil Roshankish, Sebastian Rudolph, Harald Sack, Ahmad Sakor, Jaime Salas, Thomas Schleider, Meilin Shi, Gianmarco Spinaci, Chang Sun, Tabea Tietz, Molka Tounsi Dhouib, Alessandro Umbrico, Wouter van den Berg, and Weiqin Xu. Knowledge graphs evolution and preservation – a technical report from isws 2019, 2020.

[6] Danilo Dessì, Francesco Osborne, Diego Reforgiato Recupero, Davide Buscaldi, and Enrico Motta. Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain, 2020.

[7] Zhangsheng Lai, Aik Beng Ng, Liang Ze Wong, Simon See, and Shaowei Lin. Dependently typed knowledge graphs, 2020.

[8] Bruno Sartini. Iicongraph: improved iconographic and iconological statements in knowledge graphs, 2024.

[9] Anelia Kurteva and Hélène De Ribaupierre. Interface to query and visualise definitions from a knowledge base, 2021.

[10] Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*, 56(11):13071–13102, 2023.

[11] Andrea Scharnhorst and Richard P. Smiraglia. The need for knowledge organization. introduction to the book linking knowledge: Linked open data for knowledge organization, 2022.

[12] Claude Pasquier. Biological data integration using semantic web technologies, 2009.

[13] Andi Ferhati. Clustering graphs – applying a label propagation algorithm to detect communities in graph databases, 2022.

[14] Alberto Santos, Ana R Colaço, Annelaura B Nielsen, Lili Niu, Maximilian Strauss, Philipp E Geyer, Fabian Coscia, Nicolai J Wewer Albrechtsen, Filip Mundt, Lars Juhl Jensen, et al. A knowledge graph to interpret clinical proteomics data. *Nature biotechnology*, 40(5):692–702, 2022.

[15] Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67, 2023.

[16] Juraj Vladika, Alexander Fichtl, and Florian Matthes. Diversifying knowledge enhancement of biomedical language models using adapter modules and knowledge graphs, 2023.

[17] Rita T. Sousa and Heiko Paulheim. Integrating heterogeneous gene expression data through knowledge graphs for improving diabetes prediction, 2024.

[18] Ziqi Guo, Jack Felag, Jordan C. Rozum, Rion Brattig Correia, and Luis M. Rocha. Selecting focused digital cohorts from social media using the metric backbone of biomedical knowledge graphs, 2024.

[19] Jens Dörpinghaus, Alexander Apke, Vanessa Lage-Rupprecht, and Andreas Stefan. Data exploration and validation on dense knowledge graphs for biomedical research, 2019.

[20] Amedeo Pachera, Mattia Palmiotto, Angela Bonifati, and Andrea Mauri. What if: Causal analysis with graph databases, 2024.

[21] Nandana Mihindukulasooriya, Mike Sava, Gaetano Rossiello, Md Faisal Mahbub Chowdhury, Irene Yachbes, Aditya Gidh, Jillian Duckwitz, Kovit Nisar, Michael Santos, and Alfio Gliozzo. Knowledge graph induction enabling recommending and trend analysis: A corporate research community use case, 2022.

[22] A. K. M. Bahalul Haque, B. M. Arifuzzaman, Sayed Abu Noman Siddik, Abul Kalam, Tabassum Sadia Shahjahan, T. S. Saleena, Morshed Alam, Md. Rabiul Islam, Foyez Ahmmed, and 5and Md. Jamal Hossain. Semantic web in healthcare: A systematic literature review of application, research gap, and future research avenues, 2022.

[23] Yuanyuan Tian. The world of graph databases from an industry perspective, 2022.

[24] Hassan S. Al Khatib, Subash Neupane, Harish Kumar Manchukonda, Noorbakhsh Amiri Golilarz, Sudip Mittal, Amin Amirlatifi, and Shahram Rahimi. Patient-centric knowledge graphs: A survey of current methods, challenges, and applications, 2024.

[25] Paul Cuddihy, Justin McHugh, Jenny Weisenberg Williams, Varish Mulwad, and Kareem S. Aggour. Semtk: An ontology-first, open source semantic toolkit for managing and querying knowledge graphs, 2018.

[26] V. Kiran Kumar. Semantic web approach towards interoperability and privacy issues in social networks, 2014.

[27] Rivindu Perera, Parma Nand, Boris Bacic, Wen-Hsin Yang, Kazuhiro Seki, and Radek Burget. Semantic web today: From oil rigs to panama papers, 2017.

[28] Gongjin Lan, Ting Liu, Xu Wang, Xueli Pan, and Zhisheng Huang. A semantic web technology index, 2022.

[29] Daniel Dunbar, Thomas Hagedorn, Mark Blackburn, John Dzielski, Steven Hespelt, Benjamin Kruse, Dinesh Verma, and Zhongyuan Yu. Driving digital engineering integration and interoperability through semantic integration of models with ontologies, 2022.

[30] Renzo Angles, Georg Gottlob, Aleksandar Pavlovic, Reinhard Pichler, and Emanuel Sallinger. Sparqlog: A system for efficient evaluation of sparql 1.1 queries via datalog [experiment, analysis and benchmark], 2023.

[31] Mohammed Maree and Mohammed Belkhatir. Coupling semantic and statistical techniques for dynamically enriching web ontologies, 2020.

[32] Natanael Arndt and Michael Martin. Decentralized evolution and consolidation of rdf graphs, 2019.

[33] Sydney Anuyah, Victor Bolade, and Oluwatosin Agbaakin. Understanding graph databases: A comprehensive tutorial and survey, 2024.

[34] Fabien Gandon, Franck Michel, Olivier Corby, Michel Buffa, Andrea Tettamanzi, Catherine Faron Zucker, Elena Cabrio, and Serena Villata. Graph data on the web: extend the pivot, don't reinvent the wheel, 2019.

[35] Maciej Besta, Robert Gerstenberger, Emanuel Peter, Marc Fischer, Michał Podstawski, Claude Barthels, Gustavo Alonso, and Torsten Hoefler. Demystifying graph databases: Analysis and taxonomy of data organization, system designs, and graph queries, 2023.

[36] Shiladitya Munshi, Ayan Chakraborty, and Debajyoti Mukhopadhyay. Integrating rdf into hypergraph-graph (hg(2)) data structure, 2013.

[37] Malcolm Crowe and Fritz Laux. Database technology evolution iii: Knowledge graphs and linked data, 2024.

[38] Sergio Abriola, Santiago Cifuentes, Nina Pardal, and Edwin Pin. Data-graph repairs: the preferred approach, 2023.

[39] Mohammad Mustafa Taye. Understanding semantic web and ontologies: Theory and applications, 2010.

[40] Angela Bonifati, Irena Holubová, Arnau Prat-Pérez, and Sherif Sakr. Graph generators: State of the art and open challenges, 2020.

[41] Stephen Bonner, Ian P Barrett, Cheng Ye, Rowan Swiers, Ola Engkvist, Andreas Bender, Charles Tapley Hoyt, and William L Hamilton. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. *Briefings in Bioinformatics*, 23(6):bbac404, 2022.

[42] Fritz Laux and Malcolm Crowe. Information integration using the typed graph model, 2021.

[43] Sara Hosseinzadeh Kassani and Ralph Deters. Leveraging protection and efficiency of query answering in heterogenous rdf data using blockchain, 2019.

[44] Ferdinando Villa, Stefano Balbi, Ioannis N Athanasiadis, and Caterina Caracciolo. Semantics for interoperability of distributed data and models: Foundations for better-connected information. *F1000Research*, 6:686, 2017.

[45] Samaneh Jozashoori and Maria-Esther Vidal. Mapsdi: A scaled-up semantic data integration framework for knowledge graph creation, 2019.

[46] Yuanyuan Liang, Keren Tan, Tingyu Xie, Wenbiao Tao, Siyuan Wang, Yunshi Lan, and Weining Qian. Aligning large language models to a domain-specific graph database for nl2gql, 2024.

[47] Stephen Bonner, Ufuk Kirik, Ola Engkvist, Jian Tang, and Ian P Barrett. Implications of topological imbalance for representation learning on biomedical knowledge graphs, 2022.

[48] Qiang Sun, Yuanyi Luo, Wenxiao Zhang, Sirui Li, Jichunyang Li, Kai Niu, Xiangrui Kong, and Wei Liu. Docs2kg: Unified knowledge graph construction from heterogeneous documents assisted by large language models, 2024.

[49] Shengtian Sang, Zhihao Yang, Lei Wang, Xiaoxia Liu, Hongfei Lin, and Jian Wang. Sematyp: a knowledge graph based literature mining method for drug discovery. *BMC bioinformatics*, 19:1–11, 2018.

[50] Michelle M Li, Kexin Huang, and Marinka Zitnik. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, 6(12):1353–1369, 2022.

[51] Alberto Cattaneo, Stephen Bonner, Thomas Martynec, Carlo Luschi, Ian P Barrett, and Daniel Justus. The role of graph topology in the performance of biomedical knowledge graph completion models, 2024.

[52] Chang Sun. Knowledge graph for microdata of statistics netherlands, 2021.

[53] Mona Alshahrani, Mohammed Asif Khan, Omar Maddouri, Akira R Kinjo, Núria Queralt-Rosinach, and Robert Hoehndorf. Neuro-symbolic representation learning on biological knowledge graphs, 2016.

[54] Meng Wang, Jiaheng Zhang, Jun Liu, Wei Hu, Sen Wang, Xue Li, and Wenqiang Liu. Pdd graph: Bridging electronic medical records and biomedical knowledge graphs via entity linking, 2017.

[55] Elizabeth D. Trippe, Jacob B. Aguilar, Yi H. Yan, Mustafa V. Nural, Jessica A. Brady, Mehdi Assefi, Saeid Safaei, Mehdi Allahyari, Seyedamin Pouriyeh, Mary R. Galinski, Jessica C. Kissinger, and Juan B. Gutierrez. A vision for health informatics: Introducing the sked framework.an extensible architecture for scientific knowledge extraction from data, 2017.

[56] Lauren Nicole DeLong, Ramon Fernández Mir, Zonglin Ji, Fiona Niamh Coulter Smith, and Jacques D. Fleuriot. Neurosymbolic ai for reasoning on biomedical knowledge graphs, 2023.

[57] Tommaso Soru, Diego Esteves, Edgard Marx, and Axel-Cyrille Ngonga Ngomo. Mandolin: A knowledge discovery framework for the web of data, 2017.

[58] Eric Leclercq and Marinette Savonnet. Scientific collaborations: principles of wikibridge design, 2010.

[59] Aryo Pradipta Gema, Dominik Grabarczyk, Wolf De Wulf, Piyush Borole, Javier Antonio Alfaro, Pasquale Minervini, Antonio Vergari, and Ajitha Rajan. Knowledge graph embeddings in the biomedical domain: Are they useful? a look at link prediction, rule learning, and downstream polypharmacy tasks, 2023.

[60] Partha Sen and Sumana Sen. Graph database while computationally efficient filters out quickly the esg integrated equities in investment management, 2024.

[61] Mattia Fumagalli, Marco Boffo, Daqian Shi, Mayukh Bagchi, and Fausto Giunchiglia. Towards a gateway for knowledge graph schemas collection, analysis, and embedding, 2023.

[62] Yijia Xiao, Dylan Steinecke, Alexander Russell Pelletier, Yushi Bai, Peipei Ping, and Wei Wang. Know2bio: A comprehensive dual-view benchmark for evolving biomedical knowledge graphs, 2023.

[63] Alex Averbuch and Martin Neumann. Partitioning graph databases - a quantitative evaluation, 2013.

[64] Zhao Zhang, Fuzhen Zhuang, Meng Qu, Fen Lin, and Qing He. Knowledge graph embedding with hierarchical relation structure. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3198–3207, 2018.

[65] Zhiguang Fan, Yuedong Yang, Mingyuan Xu, and Hongming Chen. Node-based knowledge graph contrastive learning for medical relationship prediction, 2023.

[66] Shusaku Egami, Takanori Ugai, Masateru Oota, Kyoumoto Matsushita, Takahiro Kawamura, Kouji Kozaki, and Ken Fukuda. Rdf-star2vec: Rdf-star graph embeddings for data mining, 2023.

[67] Yushan Liu, Marcel Hildebrandt, Mitchell Joblin, Martin Ringsquandl, Rime Raissouni, and Volker Tresp. Neural multi-hop reasoning with logical rules on biomedical knowledge graphs. In *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18*, pages 375–391. Springer, 2021.

[68] Yuanfei Dai, Shiping Wang, Neal N Xiong, and Wenzhong Guo. A survey on knowledge graph embedding: Approaches, applications and benchmarks. *Electronics*, 9(5):750, 2020.

[69] Will Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. Embedding logical queries on knowledge graphs. *Advances in neural information processing systems*, 31, 2018.

[70] Ayoub Harnoune, Maryem Rhanoui, Mounia Mikram, Siham Yousfi, Zineb Elkaimbillah, and Bouchra El Asri. Bert based clinical knowledge extraction for biomedical knowledge graph construction and analysis, 2023.

[71] Yushan Liu, Marcel Hildebrandt, Mitchell Joblin, Martin Ringsquandl, Rime Raissouni, and Volker Tresp. Neural multi-hop reasoning with logical rules on biomedical knowledge graphs, 2021.

[72] Jens Dörpinghaus, Andreas Stefan, Bruce Schultz, and Marc Jacobs. Towards context in large scale biomedical knowledge graphs, 2020.

[73] Jixue Liu, Selasi Kwashie, Jiuyong Li, Lin Liu, and Michael Bewong. Linking graph entities with multiplicity and provenance, 2019.

[74] Sendong Zhao, Bing Qin, Ting Liu, and Fei Wang. Biomedical knowledge graph refinement with embedding and logic rules, 2020.

[75] Sebastian Lobentanzer, Patrick Aloy, Jan Baumbach, Balazs Bohar, Pornpimol Charoentong, Katharina Danhauser, Tunca Doğan, Johann Dreo, Ian Dunham, Adrià Fernandez-Torras, Benjamin M. Gyori, Michael Hartung, Charles Tapley Hoyt, Christoph Klein, Tamas Korcsmaros, Andreas Maier, Matthias Mann, David Ochoa, Elena Pareja-Lorente, Ferdinand Popp, Martin Preusse, Niklas Probul, Benno Schwikowski, Bünyamin Sen, Maximilian T. Strauss, Denes Turei, Erva Ulusoy, Judith Andrea Heidrun Wodke, and Julio Saez-Rodriguez. Democratising knowledge representation with biocypher, 2023.

[76] Rahul Nadkarni, David Wadden, Iz Beltagy, Noah A. Smith, Hannaneh Hajishirzi, and Tom Hope. Scientific language models for biomedical knowledge base completion: An empirical study, 2021.

[77] Arijit Khan. Knowledge graphs querying, 2023.

[78] Fausto Giunchiglia, Alessio Zamboni, Mayukh Bagchi, and Simone Bocca. Stratified data integration, 2021.

[79] Marcelo Arenas, Claudio Gutierrez, and Juan F. Sequeda. Querying in the age of graph databases and knowledge graphs, 2021.

[80] Jens Dörpinghaus and Andreas Stefan. Optimization of retrieval algorithms on large scale knowledge graphs, 2020.

[81] Amir H. Asiaee, Prashant Doshi, Todd Minning, Satya Sahoo, Priti Parikh, Amit Sheth, and Rick L. Tarleton. From questions to effective answers: On the utility of knowledge-driven querying systems for life sciences data, 2012.

[82] Catherine Canevet, Artem Lysenko, Andrea Splendiani, Matthew Pocock, and Christopher Rawlings. Analysis and visualisation of rdf resources in ondex, 2010.

[83] Eleni Ilkou. Personal knowledge graphs: Use cases in e-learning platforms, 2022.

[84] Elisa S. Menendez, Marco A. Casanova, Mohand Boughanem, and Luiz André P. Paes Leme. Ranking rdf instances in degree-decoupled rdf graphs, 2018.

[85] Konstantinos Sikelis, George E Tsekouras, and Konstantinos I Kotis. Ontology-based feature selection: A survey, 2021.

[86] Dietmar Seipel. Knowledge engineering for hybrid deductive databases, 2017.

[87] Fritz Laux. The typed graph model – a supermodel for model management and data integration, 2021.

[88] Marko A. Rodriguez. Interpretations of the web of data, 2009.

[89] Zihao Zhao, Xiaodong Ge, and Zhihong Shen. S2ctrans: Building a bridge from sparql to cypher, 2023.

[90] Markus Schröder, Michael Schulze, Christian Jilek, and Andreas Dengel. Bridging the technology gap between industry and semantic web: Generating databases and server code from rdf, 2020.

[91] Georgios M. Santipantakis, George A. Vouros, and Christos Doulkeridis. Towards integrated and open covid-19 data, 2020.

[92] Sumit Neelam, Udit Sharma, Sumit Bhatia, Hima Karanam, Ankita Likhyani, Ibrahim Abdelaziz, Achille Fokoue, and L. V. Subramaniam. Expressive reasoning graph store: A unified framework for managing rdf and property graph databases, 2022.

[93] Jinha Kim, Hyungyu Shin, Wook-Shin Han, Sungpack Hong, and Hassan Chafi. Taming subgraph isomorphism for rdf query processing, 2015.

[94] Renzo Angles, Harsh Thakkar, and Dominik Tomaszuk. Directly mapping rdf databases to property graph databases, 2020.

[95] Nikolaos Karalis, Alexander Bigerl, and Axel-Cyrille Ngonga Ngomo. Native execution of graphql queries over rdf graphs using multi-way joins, 2024.

[96] HyeongSik Kim and Kemafor Anyanwu. Scalable ontological query processing over semantically integrated life science datasets using mapreduce, 2016.

[97] Noreddine Gherabi, Khaoula Addakiri, and Mohamed Bahaj. Discovering new technique for mapping relational database based on semantic web technology, 2021.

[98] Nikos Bikakis, Chrisa Tsinaraki, Ioannis Stavrakantonakis, Nektarios Gioldasis, and Stavros Christodoulakis. The sparql2xquery interoperability framework. utilizing schema mapping, schema transformation and query translation to integrate xml and the semantic web, 2014.

[99] Renzo Angles, Marcelo Arenas, Pablo Barceló, Aidan Hogan, Juan Reutter, and Domagoj Vrgoč. Foundations of modern query languages for graph databases. *ACM Computing Surveys (CSUR)*, 50(5):1–40, 2017.

[100] Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu, Jimmy Lin, and M Tamer Özsu. The ubiquity of large graphs and surprising challenges of graph processing. *Proceedings of the VLDB Endowment*, 11(4):420–431, 2017.

[101] Samaneh Jozashoori, Tatiana Novikova, and Maria-Esther Vidal. Data integration for supporting biomedical knowledge graph creation at large-scale, 2018.

[102] Maryam Alizadeh, Maliheh Heydarpour Shahrezaei, and Farajollah Tahernezhad-Javazm. Ontology based information integration: A survey, 2019.

[103] Ali Ben Ammar. Query optimization techniques in graph databases, 2016.

[104] Senthil Kumar Narayanasamy, Kathiravan Srinivasan, Yuh-Chung Hu, Satish Kumar Masilamani, and Kuo-Yi Huang. A contemporary review on utilizing semantic web technologies in healthcare, virtual communities, and ontology-based information processing systems. *Electronics*, 11(3):453, 2022.

[105] Andrea Splendiani, Chris J Rawlings, Shao-Chih Kuo, Robert Stevens, and Phillip Lord. Lost in translation: data integration tools meet the semantic web (experiences from the ondex project), 2011.

[106] Jaroslav Pokorny. Integration of relational and graph databases functionally, 2018.

[107] Hamed Hassanzadeh and Mohammad Reza Keyvanpour. Semantic web requirements through web mining techniques, 2012.

[108] Jiaoyan Chen, Olga Mashkova, Fernando Zhapa-Camacho, Robert Hoehndorf, Yuan He, and Ian Horrocks. Ontology embedding: A survey of methods, applications and resources, 2025.

[109] Michael E. Cotterell and Terrance Medina. A markov model for ontology alignment, 2013.

[110] Iztok Savnik, Kiyoshi Nitta, Riste Skrekovski, and Nikolaus Augsten. Statistics of knowledge graphs based on the conceptual schema, 2021.

[111] Jens Dörpinghaus, Vera Weil, Carsten Düing, and Martin W. Sommer. Centrality measures in multi-layer knowledge graphs, 2022.

[112] M. Maree and M. Belkhatir. On the merging of domain-specific heterogeneous ontologies using wordnet and web pattern-based queries, 2020.

**Disclaimer:**

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.





AI-generated, for reference only.