



A Survey of Advanced Information Fusion System: from Model-Driven to Knowledge-Enabled

Di Zhu¹ · Hailian Yin² · Yidan Xu¹ · Jiaqi Wu¹ · Bowen Zhang¹ · Yaqi Cheng¹ · Zhanzuo Yin¹ · Ziqiang Yu³ · Hao Wen⁴ · Bohan Li¹

Received: 27 September 2022 / Revised: 13 March 2023 / Accepted: 20 March 2023
© The Author(s) 2023

Abstract

Advanced knowledge engineering (KE), represented by knowledge graph (KG), drives the development of various fields and engineering technologies and provides various knowledge fusion and knowledge empowerment interfaces. At the same time, advanced system engineering (SE) takes model-based system engineering (MBSE) as the core to realize formal modeling and process analysis of the whole system. The two complement each other and are the key technologies for the transition from 2.0 to 3.0 in the era of artificial intelligence and the transition from perceptual intelligence to cognitive intelligence. This survey summarizes an advanced information fusion system, from model-driven to knowledge-enabled. Firstly, the concept, representative methods, key technologies and application fields of model-driven system engineering are introduced. Then, it introduces the concept of knowledge-driven knowledge engineering, summarizes the architecture and construction methods of advanced knowledge engineering and summarizes the application fields. Finally, the combination of advanced information fusion systems, development opportunities and challenges are discussed.

Keywords System engineering · Knowledge engineering · Knowledge graph · Model-based system engineering · Information system

1 Introduction

System engineering (SE), also known as optimization technology and simulation technology, builds a cross-domain-oriented full life cycle development framework based on

the model. Model-based system engineering (MBSE) [1] was first proposed by INCOSE in 2007. It solves the overall development problem through formal modeling throughout the life cycle. On the other hand, knowledge engineering (KE) takes knowledge as the core, expresses knowledge and

✉ Bohan Li
bhli@nuaa.edu.cn

Di Zhu
di_zhu@nuaa.edu.cn

Hailian Yin
yinhailian@nuaa.edu.cn

Yidan Xu
xuyidan2021@163.com

Jiaqi Wu
wujiaqi@nuaa.edu.cn

Bowen Zhang
727643535@qq.com

Yaqi Cheng
1421758254@qq.com

Zhanzuo Yin
1160898544@qq.com

Ziqiang Yu
zqyu@ytu.edu.cn

Hao Wen
wenhao@nuaa.edu.cn

¹ College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

² Ministry of Industry and Information Technology Key Laboratory of Advanced Technology for Small and Medium UAVs, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

³ College of Computer and Control Engineering, Yantai University, Yantai 264005, Shandong, China

⁴ College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

automatically solves problems by computer with the help of artificial intelligence. The knowledge graph (KG) was first proposed by Google in 2012 [2], trying to solve the problems of semantic expression, knowledge storage and knowledge reasoning through the graph structure.

This survey expounds the researches of SE and KE according to their research timelines and methodologies and summarizes the combination direction of the two from the perspective of double driving. The contributions of this survey are as follows:

- This survey introduces the concepts, development, methodological techniques and applications of SE and KE, respectively.
- This survey analyzes the typical representative technologies of advanced SE and advanced KE (MBSE & KG). To the best of our knowledge, MBSE can solve cross-domain data islands and achieve physical and digital mapping. At the same time, KG extracts and uses knowledge by drawing structure, which is the key enabling means of cognitive intelligence. Therefore, this survey summarizes the methods and advantages of MBSE and KG to inspire the researchers.
- Finally, this survey summarizes the technical advantages of SE and KE and analyzes the research direction, opportunities and challenges of the future development of advanced information fusion system from the perspective of dual-driver combination.

The rest of the survey is organized as follows: in Sect. 1, we systematically elaborate on the concept and development of model-driven SE, analyze the various model approaches of traditional and advanced SE deeply and summarize the key techniques and application areas of MBSE. In Sect. 2, we analyze the concept and

development of knowledge-driven KE in the same form, summarize the difference of methods between traditional KE and advanced KE in various stages and summarize the knowledge graph system architecture, construction methods, application directions and application fields. In Sect. 3, We systematically summarized the development trend of MBSE and KG two-way drive of advanced information fusion system and presented the opportunities and challenges in the future research field.

2 Model-Driven System Engineering

This chapter introduces the concepts and development of system engineering and then introduces the methods, the key technologies and the application fields of SE.

2.1 System Engineering Development

System engineering can be traced back as early as the 1940s in the field of military engineering technology. The specific development timeline is shown in Fig. 1.

(1) Application Exploration Period

In 1940, Bell Labs divided system development into five stages: planning, research, development, application and general engineering, which is the embryo of SE [3]. In 1942, the Manhattan Project used SE to coordinate the specification of atomic bomb development [4]. In 1970, NASA uses SE to send humans to the moon in the Apollo 13 project. Since then, a new era of SE development has been opened.

(2) Theoretical Formation Period

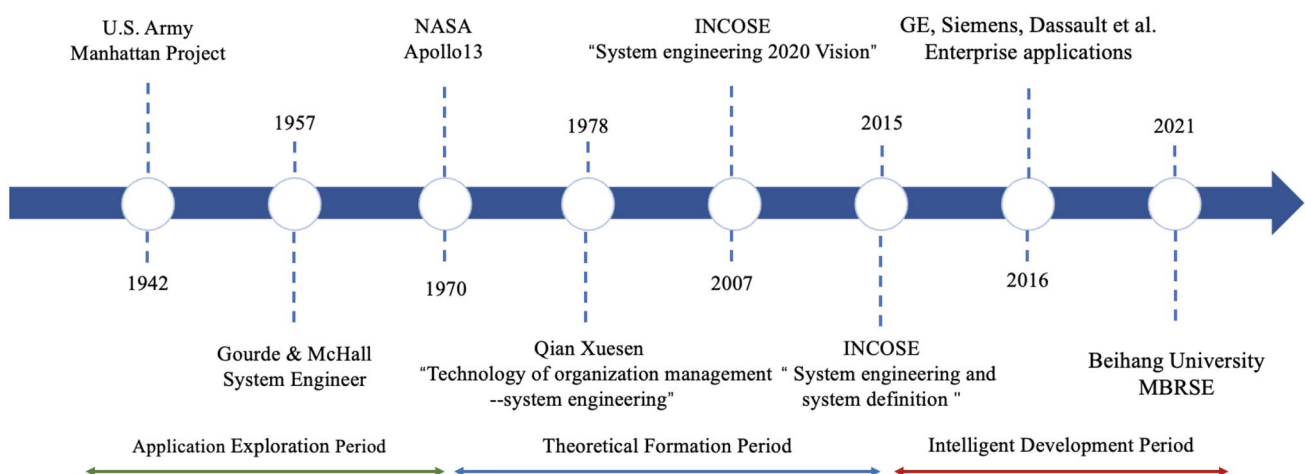


Fig. 1 System engineering development timeline

In 1957, H.H. Gourde and R.E. Mc Hall published the first SE textbook *System Engineering* [5]. In 1978, Qian Xuesen published *Technology of Organization Management - System Engineering* [6], which proposed the complete idea of SE. In 2007, INCOSE first proposed the concept of MBSE in *Vision of System Engineering 2020*. Advanced SE theory has matured.

(3) Intelligent Development Period

In 2015, INCOSE released *System and System Engineering Definitions* [7, 8], which defined SE as a way and means to make systems interdisciplinary. Since then, countries such as the USA, Germany and the UK have introduced SE into the digital transformation. Meanwhile, research institutes such as China aviation industry implemented MBSE-based digital twin applications in aircraft design [9]. In 2021, BUAA Reliability Engineering Institute established a model-based reliability system engineering (MBRSE) platform, which combines reliability design (RSE) to further develop SE theory.

2.2 Classical Methods of System Engineering

There are many methods in SE development, including traditional SE methods and advanced SE methods.

2.2.1 Document-Based System Engineering

(1) Hall Model

An SE three-dimensional analysis model is based on time dimension, logical dimension and knowledge dimension. The main methods are intuitive empirical methods, such as brainstorming (BS) [10] and synectics method [11], etc.; forecasting, such as Delphi method [12], scenario analysis [13], time series forecasting method [14], etc.; structural model method, such as interpretative structural modeling method (ISM) [15], graph theory [16], etc. The Hall model establishes the theoretical framework of DBSE. However, there are some problems such as node knowledge limitation and low portability.

(2) V-Model and V & V Model

An SE rapid development model that emphasizes model testing. Based on the waterfall model, the system design and integration process are related through model testing and developed into a V & V model [17], focusing on concurrency and risk management. Common methods include Delphi method and analytical hierarchy process (AHP). For example, Wang [18] proposed the $I + X$ education resource

evaluation framework based on Delphi method. V model and V & V model can realize visual management in complex systems. However, there are some defects such as difficult backtracking, large rework amount and low flexibility.

(3) Spiral Model

A risk-driven SE method, which carries out risk identification, risk analysis and risk control for activities in each stage in advance. Applicable to interdisciplinary complex SE, such as B The OSCEs (objective structured clinical examinations) proposed by kumaravel [19] et al., is used to evaluate the medical ability of students. Spiral model focuses on the development of prototype system, which has the advantages of flexible design and controllable interaction, but it relies heavily on risk identification.

In addition, DBSE includes feedback models, elliptical models and circular models. DBSE has been successfully applied in the space industry in the early years, but gradually unable to cope with the system complexity brought by the development of information and network. Therefore, MBSE came into being.

2.2.2 Model-Based System Engineering

(1) MBSE

A formal application of a modeling approach that supports activities such as demand, design, analysis, validation and validation through the life cycle. Compared with DBSE, it has the following advantages: 1) integrated design, covering the full life cycle, guaranteeing consistency management, traceability management and complexity management; 2) multidisciplinary collaborative validation, through simulation technology to achieve system pre-validation and interdisciplinary optimization; and 3) features are unambiguous, graphical models are used to accurately and uniformly describe system functions, standardize system design and transfer information without ambiguity. Currently, MBSE is the mainstream application and research method in large and complex systems in many fields.

(2) MBRSE

Model-based reliability system engineering (MBRSE) [20] was proposed by the School of Reliability and System Engineering, Beihang University in 2021. Based on product, fault and environment models, it cognizes and applies failure rules according to model evolution and achieves closed-loop optimization and control of failure. The deep combination of MBSE and reliability system engineering (RSE) is an important development direction of advanced system engineering in the future.

Table 1 Comparison of system engineering methods

	Traditional DBSE	Advanced MBSE
Technical features	Mechanization, electronics, automation and scale	Networking, intelligence and digital transformation
Organizational model	Specialization in specific industry areas	Value chain integration in all fields
Theoretical basis	Hall Model, Waterfall Model, V Model	MBSE, MBRSE, Digital Twin
Drive data	Natural language-based documents, etc.	Model-based on standard language
Cons/Pros	Semantic ambiguity, text static, low text relevance, passive delay of demand change, etc.	High consistency, traceability, dynamic correlation throughout the life cycle and real-time global response to demand changes
System characteristics	Electromechanical System [Number of parts + lines of source code] $< 10^6$	Cyber-Physical System [Number of parts + lines of source code] $> 10^6-10^7$
Authentication method	Test + Simulation	Modeling simulation + Cross-Domain Experiment + Cloud computing
Application scenarios	Professional integration + data island	Model-based cross-domain full life cycle management

(3) MBSE-Based Digital Twin

A data analysis framework [21] that combines MBSE with cyber-physical systems (CPSs). Based on high-precision digital system model (DSM), this framework constructs a mirror of physical entities in digital space, maps the entire life cycle process of corresponding entities and provides management, integration and analysis capabilities. It has the technical features of *all element modeling definition, all data collection and all decision simulation*. Essentially, it is to establish a digital mirror (DM) [22] of physical entities in SE.

In summary, SE can be divided into DBSE and MBSE as shown in Table 1.

2.3 Key Technologies

2.3.1 MBSE Modeling Language

Including DSL, UML, SysML, etc., [23] DSL refers to the setting language for a specific field, which has strong description ability for a specific range and solves the communication and consistency problems of MBSE. It can be divided into internal DSLs such as Kotlin DSL and groovy DSL, and external DSLs such as regular expression, XML and Markdown. SysML is a general graphical modeling language. It is reused and extended based on Unified Modeling Language (UML), which formally represents the system requirements, functions and processes in SE, and effectively solve the consistency problem of different model languages.

2.3.2 MBSE Formal Methods

A method of normalizing formal structure in SE through logic science, combining mathematical analysis and computer verification to improve reliability and robustness, is applied to highly integrated SE that pays attention to safety,

including formal specification (FS) and formal verification (FV). FS uses DSL to induce the logical relationship of the system, including process algebra language CSP, CCS, etc., and sequential logic language PLTL, CTL, etc. FV verifies whether the formalized system meets the design requirements can be verified by automatic theorem proving tools such as CoQ [24] and lean [25], or distributed verification tools based on time sequence such as UPPAAL [26], TLA [27], Spin [28, 29].

2.4 Advanced System Engineering Applications

System engineering has full life cycle data processing capabilities. International enterprises such as SIEMENS, DASSAULT, PTC, Ansys, AVEVA, Microsoft and Unity have continuously developed system engineering applications. Based on industry requirements, model-driven system engineering application systems, including GE Predix, SIEMENS COMOS, PTC ThingWorx and Jupiter Digital Twin Platform have been built [30–33]. Advanced SE has been studied and practiced in many fields and can be summarized as urban governance, industrial manufacturing and health care.

(1) Urban governance

with the increasing scale of the city, MBSE builds a city image in the virtual space using urban facilities and human behavior learning to predict the changes of the city state [34]. Optimize the comprehensive management of traffic information [35]. In addition, by building an environmental assessment model based on MBSE to detect air pollution and greenhouse gas emissions [36], we can also reduce pollution emissions and protect urban ecology.

(2) In industrial manufacturing

in the modular production process, MBSE manages each life cycle stage of manufacturing in series [37] and

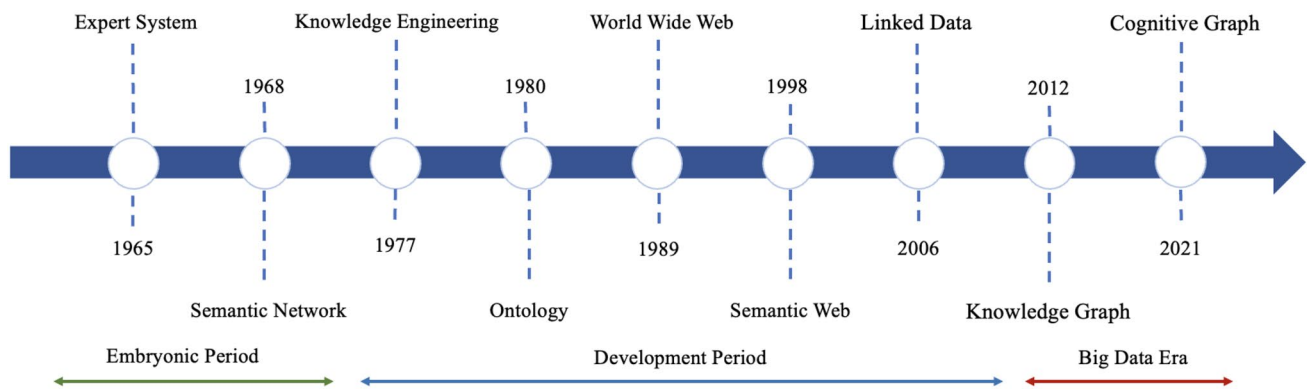


Fig. 2 Knowledge engineering development timeline

realizes the visual display of product manufacturing [38]. At the same time, MBSE supports cross-domain designer information interaction, including intelligent order scheduling, system decision support and other operations, integrates, analyzes and processes information flow and realizes data exchange more efficiently.

(3) Health Care

With the help of medical Internet of Things equipment, MBSE has established a new medical simulation method to conduct real-time perception of the patient's body data [39], track health care data such as diet, sleep and exercise, evaluate health status and give diagnosis results and treatment plans. In addition, combined with VR, AR and other technologies, digital twin technology based on MBSE makes remote surgery possible [40].

3 Knowledge-Enabled Knowledge Engineering

This chapter first introduces the development of knowledge engineering and then introduces the architecture, methods and applications of advanced KE.

3.1 Knowledge Engineering Development

Knowledge engineering can be traced back to the expert system in the last century. The knowledge graph development timeline is shown in Fig. 2.

- (1) Embryonic Period–Expert System in 1965, E. A. Feigenbaum presents expert systems (ESs) [41] for making decisions based on expert knowledge. In 1968, Quillian proposed semantic network (SN), which consists of connected nodes (concepts or objects) and edges (relationships between nodes). Since then, knowledge base

(KB) and knowledge representation (KR) have become hot spots.

- (2) Development Period–Semantic Web in 1977, E. A. Feigenbaum proposed knowledge engineering (KE) [42], which uses artificial intelligence, takes knowledge as the object of processing and uses computers to solve problems. In 1980, McCarthy proposed ontology [43], which describes the world by building ontologies. In 1989, Tim Berners-Lee invented the World Wide Web (WWW) [44], and in 1998 introduced the semantic web (SW) [45], which made network data machine-readable. After the WWW was combined resource description framework (RDF), knowledge representation and reasoning were acquired. In 2006, Berners Lee proposed linked data [46], a semantic web ecosystem for data disclosure and data linking.
- (3) Big Data Era–Knowledge Graph

In 2012, Google introduced the concept of knowledge graph to improve the return quality and query efficiency of search engines and provide structured information around topics. Since then, KG has become the key technology of advanced KE. In 2021, the Institute of Artificial Intelligence of Tsinghua University and others released the *Research Report of Cognitive Graph*, which uses KG, cognitive reasoning and logical expressions to form a cognitive graph, enabling knowledge to be understood and used by machines, realizing intelligence from perception to a critical breakthrough in mental intelligence.

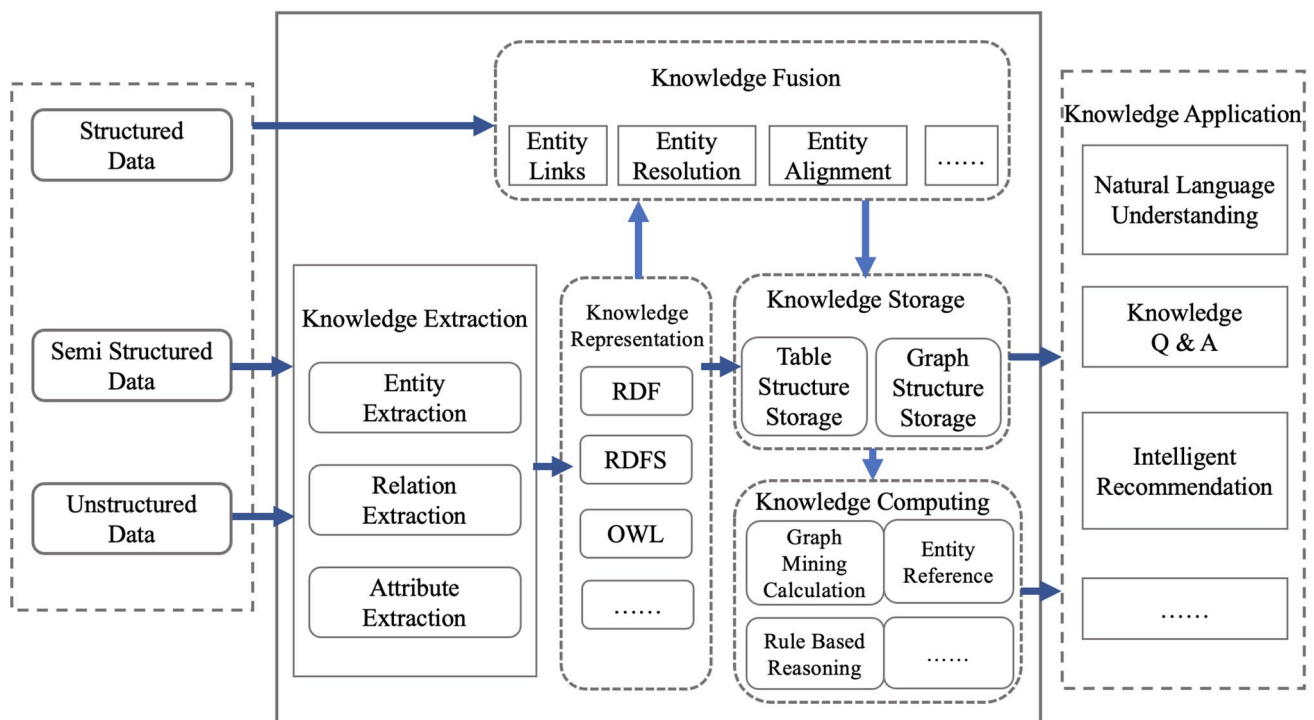
KE has mainly experienced three development stages, and the data comparison of each development stage of KE are shown in Table 2.

3.2 Advanced Knowledge Engineering Architecture

The architecture of KE is shown in Fig. 3. It first obtains and processes raw data (unstructured, semi-structured and structured), then extracts information elements through

Table 2 Comparison of development stages of knowledge engineering

	Expert system	Semantic web	Knowledge graph
Data scale	Experts build knowledge base data volume <math> < 10^5 </math>	Build knowledge base and inference engine manually data volume: <math> <="" -="" 10^4="" 10^6="" math><="" td=""> <td>Automatic construction based on machine learning, natural language processing and other methods Data volume > <math> <="" 10^9="" math><="" td=""> </math>></td></math>>	Automatic construction based on machine learning, natural language processing and other methods Data volume > <math> <="" 10^9="" math><="" td=""> </math>>
Knowledge representation	Production rules, semantic networks, logical patterns scripts, etc.	Value chain integration in all fields	RDF triples and attribute maps, etc
Application scenario	Simulate expert thinking, carry out active reasoning and judgment and solve problems	Logical reasoning, (certainly & uncertainly reasoning)	Fact knowledge retrieval and reasoning task capability

**Fig. 3** Architecture of knowledge engineering

entity extraction, relation extraction, attribute extraction and other knowledge extraction methods and performs knowledge fusion of entity links, entity resolution, entity alignment, etc., based on knowledge representation methods such as EDF, RDFs, OWL, etc., extracts knowledge facts from the original database and third-party databases, and builds knowledge base to realize knowledge computing of graph mining calculation, entity reference, rule-based reasoning, etc., and finally realizes knowledge application.

3.3 Advanced Knowledge Engineering Methods

The main methods of KE can be divided into knowledge extraction, knowledge fusion and knowledge processing.

3.3.1 Knowledge Extraction

Knowledge extraction (KE) focuses on extracting important information elements from heterogeneous data, mainly divided into named entity recognition and relationship extraction.

- (1) Name entity recognition Name entity recognition (NER) identifies entity types with specific meanings from text. The main methods include rule-based NER identifies entities by matching rules between the marked data and the natural text according to the expert rule template; Statistical-learning-based NER focuses on the problem of sequence labeling through

the semantic features of the text, and uses the model to identify after iteratively calculating the training set; deep learning-based NER is the mainstream method at present.

(2) Relation extraction

(RE), after NER, establishes semantic links between entities to form triple knowledge for discretized semantic relationships. The main methods include rule-based RE, which constructs rules manually, has the advantages of high accuracy and strong targeting, but relies heavily on manual operation and high cost; statistical-learning-based RE divided into three categories. Supervised RE acquire features through sequence markers and context relationships. Semi-supervised RE requires only a small amount of manual label data. Unsupervised RE do not require manual labeling and have a

strong migration ability; deep learning-based RE automatically extracts features and mines hidden information through a large number of training, is the main RE method at present. The common knowledge extraction methods is shown in Table 3.

3.3.2 Knowledge Fusion

Knowledge fusion (KF) transforms the obtained facts into a structured knowledge system. The main methods are as follows:

- (1) Entity Disambiguation Entity disambiguation (ED) solves the problem of polysemy, is divided into cluster

Table 3 Common knowledge extraction methods

Method	Model	Characteristics	Data sets	F1/%
NER	BILSTM-CRF [47]	Capturing past and future features depends on feature engineering	CoNLL2003	90.1%
NER	GRN [48]	Using GRN to capture long-term dependency context information	CoNLL2003	92.0%
			OntoNotes5	87.7%
NER	LGN [49]	Modeling text entity extraction problem as a graph node classification problem in graph networks	MSRA	93.4%
			Weibo NER	95.3%
NER	W2NER [50]	Modeling adjacency relationships between entity words and next adjacent and tail word relationships	CoNLL2003	93.0%
			ACE2005	86.7%
			CADEC	73.2%
ER	depLCNN [51]	Using shortest dependent path and negative sampling to avoid interference of unrelated sequences	SemEval2010	84.0%
ER	STGCN [52]	Converting a dependency tree to a weighted graph connection node	SemEval2010	86.0%
ER	ChMEL [53]	Extracting entity relationships across sentences or paragraphs from chapter level	SemEval2010	85.9%
ER	GDPNet [54]	Using GGG to generate multi-view edges, the graph is refined by DTWPoo	DialogRE	80.2%

Table 4 Common knowledge fusion methods

Method	Model	Characteristics	Data sets	F1/%
ED	DNN-ED [55]	Directly optimizing document and entity representations for a given similarity measure	TACKBP2010	94.0%
ED	MCE-ED [56]	Encoding references, contexts and entities in continuous vector space to entity disambiguate	TACKBP2009	82.3%
			TACKBP2010	83.9%
ED	LNA-ED [57]	Entity embedding, neural attention mechanism on local context window and differentiable joint reasoning stage	AIDA MSNBC	92.22%
				93.7%
ED	ExtEnd [58]	Define the task as a text extraction problem, which is solved by two transformer-based architectures	AIDA MSNBC	90.0%
				94.5%
EA	IPTransE [59]	Encoding jointly various KGs in entities and relationships into unified low-dimensional semantic space.	DFB-1	86.5%
			DFB-2	82.2%
			DFB-3	47.4%
EA	MuGNN [60]	Learning alignment-oriented KG embedding through multichannel robust coding of two KG	DBP	89.7%
EA	PREA [61]	Abstracting the existing EA methods into unified framework Shape-BuilderAlignment	DBP	92.9%
EA	SelfKG [62]	Reconsidering the use of supervision in EA to align entities with same meaning in different KGs	DBP	91.3%

disambiguation and link disambiguation. Clustering disambiguation aggregates entities with high similarity by comparing similarity. Link disambiguation is to disambiguate the corresponding target by referring to the linked entity.

- (2) Entity Alignment Entity alignment (EA) uses the attribute information of entities to determine whether different entities can be aligned and calculates the entity similarity method by mapping entities and relationships to low-dimensional vectors. The common knowledge fusion methods are shown in Table 4.

3.3.3 Knowledge Processing

Knowledge processing (KP) transforms the obtained facts into a structured knowledge system. The main methods are as follows:

- (1) Ontology Construction
Ontology construction obtains the ontological database through the similarity calculation of entity relations, the extraction of superior and subordinate relations and the generation of ontologies, and provides the definition, principle and theory of specific fields.
- (2) Knowledge Reasoning
Knowledge reasoning infers entity relationships and constructs new relationships based on entity relationship data in the knowledge base, thus expanding and enriching the knowledge base network.
- (3) Quality Evaluation

Quality evaluation screens knowledge, ABC quantifies the trustworthiness of knowledge and discards low-quality knowledge data to ensure the quality of knowledge base.

The common methods of knowledge reasoning are shown in Table 5:

Table 5 Common methods of knowledge reasoning

Reasoning classification	Main idea	Typical method
Graph structure & statistical rules	Using graph structure to mine the rules contained in KG for reasoning	SFE [63], HIRI [64], PRA [65], CPRA [66]
Representation learning	Representation of entity relations by vectors, matrices and tensors	TransE [67], RESCAL [68], DistMul [69], ANALOGY [70]
Neural network	Vectorizing fact tuples, continuously training the output scores of the neural network to select entities	InteractE [71], Convolutional 2D [72], NTN [73]
Hybrid reasoning	Reasoning by mixing the advantages of each modes	Rules & Distributed KR [74], HNN & Distributed KR [75]

3.4 Knowledge Engineering Application

Structured KE has the ability of efficient data processing and knowledge reasoning, and has been applied in various fields.

- (1) Smart finance KE stores massive data such as financial information and investment data in the form of triples, mines relationships, carries out financial recommendation, investment inquiry, customer mining and so on. At the same time, KG can realize intelligent risk identification and independent early warning.
- (2) Smart Medical

KE integrates medical data, assists doctors in decision-making and provides personalized medication services. The intelligent question and answer system based on knowledge atlas can realize intelligent inquiry. At the same time, KE can build clinical decision support (CDS) composed of KG, reasoning engine and medical experts.

- (2) Smart City

KE extracts citizen attributes, excavates service site loads and proposes optimal policies to provide effective decision-making and reference for city management. Implement traffic analysis and modeling of urban services, urban traffic management and public data mining. Moreover, the City-KG can generate dynamic management information to dynamically warn the city security and avoid potential risks.

4 Advanced Information Fusion System

This section firstly introduces the information fusion system of KE and SE and summarizes the development ideas and emerging technologies of KG for MBSE and MBSE for KG.

4.1 Information Fusion System of KE and SE

In the era of AI 3.0, perceptual intelligence is turning to cognitive intelligence. The fusion of KE and SE can effectively help the realization of knowledge empowerment.

Recently, due to the increasing requirements of information systems on data, models and knowledge, the development of advanced KE and SE has made the information fusion systems important. On the basis of retaining the original system characteristics of advanced KE and SE methods, the advanced information fusion system has been upgraded from model-driven to knowledge-enabled through a variety of fusion forms of KG and MBSE.

Advanced information fusion system was mainly applied in the aerospace field in the early stage. For example, NASA integrated digital thread with knowledge architecture and realized the domain integration of knowledge management, information architecture and data science through digital twins [76, 77]. At present, informatization and digitalization have promoted information fusion systems to many fields. Based on the original MBSE information system, SIEMENS, DASSAULT, Microsoft and other enterprises integrate KG technology, expand the range of associated data, map multisource production models and provide visualization of complex networks [33, 78].

The combination of SE and KE brings a new vision for future development. KG enables MBSE, which intelligently implements structural design, reliability design, knowledge penetration and so on. Meanwhile, MBSE offers KG formal modeling, data chain governance, digital twinning, and more from a new perspective. This chapter summarizes the idea and technology of combining KE and SE in advanced information fusion system.

4.2 KG for MBSE

MBSE data processing is based on empirical methods and domain rules. The introduction of KG provides specific solutions to the structural design, reliability and security problems of system engineering.

4.2.1 Design Method of Span Structure

MBSE relies on traditional data storage methods and has optimization space for storage, query and reasoning. The introduction of KG brings the following advantages:

- (1) **Data Integration** The existing MBSEs have weak ability to integrate the same model, different sources and different forms. KG can use GOPPRRE, OWL and other methods to regularize the pre-processing of MBSE model, build a structured knowledge map for the whole life cycle and ensure the consistency and integrity of module data. For example, Zuheros [79, 80] et al. proposed to use LSTM to encode the context information of the target entity without using external resources such as knowledge base.

- (2) **Graph Fusion MBSE** has problems such as data maintenance, redundant documents and understanding ambiguity. KG can provide RDF standard data model [81]. It can use visualization technology to describe the knowledge and relationship of MBSE, such as DB2RDF [82] and SQLGraph [83]. It takes into account the physical structure of relational tables and the logical structure of graph models [84].
- (3) **Multimodal Embedding MBSE** relies on expert experience to artificially represent data and query models, and traditional data query methods cannot meet large-scale dynamic query requirements. Deep learning-based KG can automatically learn hidden features and identify synonym models from MBSE when domain knowledge is scarce. For example, large-scale concept ontology for multimedia (LSCOM) [85–88], core ontology for multimedia (COMM) [89], etc.

4.2.2 Implementation of Highly Reliable Model

MBSE is an important research direction in SE. The introduction of KG helps the combination of MBSE and RSE in the following ways:

- (1) **Sensitive Data Supervision** MBSE's model data are stored and transmitted on a variety of device ports. Sensitive data leaking will lead to information loss and system security problems. Security knowledge map can protect MBSE's network data, establish ontology-based security knowledge ontology framework and construct structured knowledge base of smart security domain by threat modeling, such as STUCCO [90], UCO, MALOnt [91], etc.
- (2) **Failure Behavior Analysis** When MBSE conducts potential failure modes and consequence analysis (e.g., FMEA, FTA, FA, etc.), there is a problem of failure relationship split. With the help of NLP training model, failure induction KG can be constructed to achieve intelligent fault diagnosis and failure attribution analysis and reduce the loss caused by failure.
- (3) **Attack Investigation**

MBSE's single-step attack warning cannot meet the requirements of network early warning. Security reasoning KG can determine attack source, attack medium and attack path. It is an important technology for MBSE from passive to active defense, such as CVE [92], CCE [93], CVSS [94] and CAPEC [95]. At the same time, associating security knowledge map with log semantics can analyze attack path by graph, such as extracting the semantics of attack behavior proposed by Zeng J [96], inferring node semantics, enumerating all behaviors.

4.3 MBSE for KG

KG is conducting further research in several fields, and the intelligent integration with MBSE provides specific ideas for the development of cognitive intelligence.

4.3.1 Data Chain Governance

Accurate extraction and rapid aggregation of KG knowledge relies on high-quality data, and MBSE offers the following data chain governance advantages:

- (1) Discovery of All Cycle Data KG automatically finds datasets from a data warehouse (DWH). Combined with MBSE, it provides data integration and description for each stage. At the same time, ML-based data discovery can further enhance the ability of data query [97–99]. For example, Aurum proposed by Fernandez et al., based on hyper-graphic data discovery system, provides queries for search datasets according to user needs.
- (2) Cleanup of All System Data KG does not have a good way to handle dirty read. MBSE provides high-quality data by detecting and repairing periodic datasets in advance. For example, an active cleaning framework based on machine learning proposed by Wang [100] et al., selects records that maximize model performance and iteratively processes them.
- (3) Annotation of All Domain Data

KG relies heavily on data quality in model training. MBSE can provide high-quality labeling services, which can tag a large amount of training data for KG using experts in various fields, industry knowledge base and enterprise crowdsourcing, for example, platform [101] such as Mechanical Turk, Figure-eight, Crowd-Flower, Mighty AI, or domestic Baidu public test, Ali Crowdsourcing, etc.

4.3.2 Digital Twin Management

KG lacks complete cross-domain engineering management capabilities. MBSE-based digital twins (DT) provides a grip for accelerating KG applications.

- (1) DT-based System Analysis Systems analysis of KG applications is incomplete, and the introduction of DT can improve interpretability. In the digital simulation system, cluster method, factor analysis, co-word analysis is processed, and the feature vector of the knowledge map is centrally sampled to make the model explanatory. Finally, the model is further understood in the form of visualization.
- (2) DT-based Security Perception

KG relies on traditional security detection methods and has insufficient processing power to determine hidden threats, dynamic threat assessment, threat event mining and so on. MBSE-based DT can effectively analyze data and monitoring systems and build a knowledge base of KG's heterogeneous cyber security information (HCSI) intelligent security domain to achieve visual security awareness and detection.

5 Conclusion

Advanced SE takes MBSE as its core and can process large data throughout its life cycle. Meanwhile, advanced KE takes KG as its representative and is the key enabling tool for cognitive intelligence. Both SE and KE are looking for new interdisciplinary possibilities. This survey reviews the research and application of SE and KE and introduces their concepts, development, methods and applications, respectively. In addition, we analyze and discuss the development form of advanced information fusion system from model-driven to knowledge-enabled. Hoping this review can provide theoretical references and innovative ideas for the research and development of advanced information fusion system.

Acknowledgements Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor Bohan Li, whose expertise was invaluable in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I would particularly like to acknowledge my teammate, for their wonderful collaboration and patient support. I would also like to thank my tutors Bohan Li, for their valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and successfully complete my dissertation. In addition, I would like to thank my parents for their wise counsel and sympathetic ear. You are always there for me. Finally, I could not have completed this dissertation without the support of my friends, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research

Author Contributions DZ gave his contribution to the collation and analysis of references and the compilation of manuscripts; BL provided the idea and gave analysis through constructive discussions; HY, YX, JW, BZ updated the references in the digital thread section; YC and ZY updates the references in the map of knowledge section; HW and ZY summarized the information fusion system section.

Funding 1. The National Key R &D Program of China under Grant (2020YFB1708100). 2. National Natural Science Foundation of China (62,172,351). 3. 14th Five-Year Plan Civil Aerospace Pre-research Project of China (D020101). 4. the Fund of Prospective Layout of Scientific Research for NUAA(Nanjing University of Aeronautics and Astronautics).

Availability of Data and Materials All data generated or analyzed during this survey are included in this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Friedenthal S, Griego R, Sampson M (2007) IncoSE model based systems engineering (mbse) initiative. In: INCOSE 2007 symposium, vol. 11
- Singhal A et al (2012) Introducing the knowledge graph: things, not strings. *Off Google Blog* 5:16
- Schlager KJ (1956) Systems engineering—key to modern development. *IRE Trans Eng Manag* 3:64–66
- Hao Y, FJH (2007) Systems engineering methods and applications
- Good H, Machol R (1957) Systems engineering. An introduction to the design of large-scale systems
- Qian XS, WSY, Xu GZ, (2011) The technology of organization management: systems engineering. *J Univ Shanghai Sci Technol* 6:520–525
- INCOSE (2013) Systems engineering manual: guide to system lifecycle processes and activities
- Walden DD, Roedler GJ, Forsberg K (2015) IncoSE systems engineering handbook version 4: updating the reference for practitioners. *INCOSE Int Symp* 25:678–686
- Hause M (2017) Using mbse to evaluate and protect the electrical grid as a system of systems. *INCOSE Int Symp* 27:597–612
- Stroebe W, Nijstad BA, Rietzschel EF (2010) Beyond productivity loss in brainstorming groups: the evolution of a question. *Adv Exp Soc Psychol* 43:157–203
- Aiamy M, Haghani F (2012) The effect of synectics brainstorming on 3rd grade students' development of creative thinking on science. *Procedia Soc Behav Sci* 47:610–613
- Laakso K, Rubin A, Linturi H (2010) Delphi method analysis: the role of regulation in the mobile operator business in Finland. In: *PICMET 2010 technology management for global economic growth*, pp 1–7
- Meixell MJ, Wu SD (2001) Scenario analysis of demand in a technology market using leading indicators. *IEEE Trans Semicond Manuf* 14(1):65–75
- Kocak C, Egrioglu E, Bas E (2021) A new deep intuitionistic fuzzy time series forecasting method based on long short-term memory. *J Supercomput* 77(6):6178–6196
- Zaidi SAH, Shahbaz M, Hou F, Abbas Q (2021) Sustainability challenges in public health sector procurement: an application of interpretative structural modelling. *Socioecon Plann Sci* 77:101028
- Chen X, Wang HH, Tian B (2019) Visualization model of big data based on self-organizing feature map neural network and graphic theory for smart cities. *Clust Comput* 22(6):13293–13305
- Cotterman H, Forsberg K, Mooz H (2005) Visualizing project management: models and frameworks for mastering complex systems
- Wang X, Chen T, Zhang Y, Yang HH (2021) Implications of the delphi method in the evaluation of sustainability open education resource repositories. *Educ Inf Technol* 26(4):3825–3844
- Kumaravel B, Stewart C, Ilic D (2021) Development and evaluation of a spiral model of assessing ebm competency using osces in undergraduate medical education. *BMC Med Educ* 21(1):1–9
- Ren Y, Wang YDZZL (2021) Model based reliability systems engineering
- Tao J, Dai WR, (2017) Study on production lifecycle based on digital thread and digital twin. *Aeronaut Manuf Technol* 21(6):26–31
- Grieves M (2014) Digital twin: manufacturing excellence through virtual factory replication. *White paper* 1(2014):1–7
- Fowler M (2010) Domain-specific Languages
- Barras B, Boutin S, Cornes C, Courant J, Coscoy Y, Delahaye D, de Rauglaudre D, Filiâtre J-C, Giménez E, Herbelin H et al (1999) The coq proof assistant reference manual. INRIA, version 6(11)
- Moura Ld, Kong S, Avigad J, Doorn Fv, Raumer Jv (2015) The lean theorem prover (system description). In: *International conference on automated deduction*, pp 378–388. Springer
- Behrmann G, David A, Larsen KG (2004) A tutorial on uppaal. *Formal methods for the design of real-time systems*, pp 200–236
- Lamport L (1999) Specifying concurrent systems with tla+. *Calculational System Design*, pp 183–247
- Holzmann GJ (2004) Book review: The spin model checker—primer and reference manual 21(2):143–151
- Cimatti A, Clarke E, Giunchiglia F, Roveri M (2000) Nusmv: a new symbolic model checker. *Int J Softw Tools Technol Transfer* 2(4):410–425
- Chen Y (2017) Integrated and intelligent manufacturing: perspectives and enablers. *Engineering* 3(5):588–595
- Scheeren I, Pereira CE (2014) Combining model-based systems engineering, simulation and domain engineering in the development of industrial automation systems: Industrial case study. In: *2014 IEEE 17th International symposium on object/component/service-oriented real-time distributed computing*, pp 40–47. IEEE
- Singh S, Shehab E, Higgins N, Fowler K, Erkoyuncu JA, Gadd P (2021) Towards information management framework for digital twin in aircraft manufacturing. *Procedia CIRP* 96:163–168
- Lim KYH, Zheng P, Chen C-H, Huang L (2020) A digital twin-enhanced system for engineering product family design and optimization. *J Manuf Syst* 57:82–93
- Cimatti A, Clarke E, Giunchiglia F, Roveri M (2000) Nusmv: a new symbolic model checker. *Int J Softw Tools Technol Transfer* 2(4):410–425
- Khan A, Aslam S, Aurangzeb K, Alhusein M, Javaid N (2022) Multiscale modeling in smart cities: a survey on applications, current trends, and challenges. *Sustain Cities Soc* 78:103517
- Oxley T, Dore AJ, ApSimon H, Hall J, Kryza M (2013) Modelling future impacts of air pollution using the multi-scale uk integrated assessment model (ukiam). *Environ Int* 61:17–35
- Rosen R, Von Wichert G, Lo G, Bettenhausen KD (2015) About the importance of autonomy and digital twins for the future of manufacturing. *IFAC papersonline* 48(3):567–572
- Kritzinger W, Karner M, Traar G, Henjes J, Sihn W (2018) Digital twin in manufacturing: a categorical literature review and classification. *IFAC PapersOnLine* 51(11):1016–1022
- Viola J, Chen Y (2021) Parallel self optimizing control framework for digital twin enabled smart control engineering. In: *2021 IEEE 1st International conference on digital twins and parallel intelligence (DTPI)*, pp 358–361

40. Laaki H, Miche Y, Tammi K (2019) Prototyping a digital twin for real time remote control over mobile networks: application of remote surgery. *Ieee Access* 7:20325–20336
41. Liao S-H (2005) Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Syst Appl* 28(1):93–103
42. Studer R, Benjamins VR, Fensel D (1998) Knowledge engineering: principles and methods. *Data Knowl Eng* 25(1–2):161–197
43. Deng ZH, TSW, (2002) Survey of ontology research. *Acta Sci Nat Univ Pekin* 5(9):730–738
44. Berners Lee T, Cailliau R, Groff JF, Pollermann B (1992) World-wide web: the information universe. *Int Res* 2(1):52–8
45. Shadbolt N, Berners-Lee T, Hall W (2006) The semantic web revisited. *IEEE Intell Syst* 21(3):96–101
46. Peng Y, Choi B, Xu J (2021) Graph learning for combinatorial optimization: a survey of state-of-the-art. *Data Sci Eng* 6(2):119–141
47. Huang Z, Xu W, Yu K (2015) Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*
48. Chen H, Lin Z, Ding G, Lou J, Zhang Y, Karlsson B (2019) Grn: Gated relation network to enhance convolutional neural network for named entity recognition. *Proc AAAI Conf Artif Intell* 33:6236–6243
49. Gui T, Zou Y, Zhang Q, Peng M, Fu J, Wei Z, Huang X-J (2019) A lexicon-based graph neural network for chinese ner. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp 1040–1050
50. Li J, Fei H, Liu J, Wu S, Zhang M, Teng C, Ji D, Li F (2022) Unified named entity recognition as word-word relation classification. *Proc AAAI Conf Artif Intell* 36:10965–10973
51. Xu K, Feng Y, Huang S, Zhao D (2015) Semantic relation classification via convolutional neural networks with simple negative sampling. *arXiv preprint arXiv:1506.07650*
52. Sun K, Zhang R, Mao Y, Mensah S, Liu X (2020) Relation extraction with convolutional network over learnable syntax-transport graph. *Proc AAAI Conf Artif Intell* 34:8928–8935
53. Lin HL, Wang YZ, Jia YR (2017) Network big data oriented knowledge fusion methods: a survey. *Chinese J Comput* 40(1):1–27
54. Xue F, Sun A, Zhang H, Chng ES (2021) Gdpnet: refining latent multi-view graph for relation extraction. *Proc AAAI Conf Artif Intell* 35:14194–14202
55. He Z, Liu S, Li M, Zhou M, Zhang L, Wang H (2013) Learning entity representation for entity disambiguation. In: Proceedings of the 51st annual meeting of the association for computational linguistics Vol 2, pp 30–34
56. Sun Y, Lin L, Tang D, Yang N, Ji Z, Wang X (2015) Modeling mention, context and entity with neural networks for entity disambiguation. In: Twenty-fourth International joint conference on artificial intelligence
57. Ganea OE, Hofmann T (2017) Deep joint entity disambiguation with local neural attention. *arXiv preprint arXiv:1704.04920*
58. Barba E, Procopio L, Navigli R (2022) Extend: extractive entity disambiguation. In: Proceedings of the 60th annual meeting of the association for computational linguistics Vol 1, pp 2478–2488
59. Zhu H, Xie R, Liu Z, Sun M (2017) Iterative entity alignment via knowledge embeddings. In: Proceedings of the international joint conference on artificial intelligence (IJCAI)
60. Cao Y, Liu Z, Li C, Li J, Chua T-S (2019) Multi-channel graph neural network for entity alignment. *arXiv preprint arXiv:1908.09898*
61. Mao X, Wang W, Xu H, Wu Y, Lan M (2020) Relational reflection entity alignment. In: Proceedings of the 29th ACM international conference on information knowledge management, pp 1095–1104
62. Liu X, Hong H, Wang X, Chen Z, Kharlamov E, Dong Y, Tang J (2022) Selfkg: self-supervised entity alignment in knowledge graphs. In: Proceedings of the ACM web conference 2022, pp 860–870
63. Gardner M, Mitchell T (2015) Efficient and expressive knowledge base completion using subgraph feature extraction. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 1488–1498
64. Liu Q, Jiang L, Han M, Liu Y, Qin Z (2016) Hierarchical random walk inference in knowledge graphs. In: Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval, pp 445–454
65. Wang Q, Liu J, Luo Y, Wang B, Lin C-Y (2016) Knowledge base completion via coupled path ranking. In: Proceedings of the 54th annual meeting of the association for computational linguistics, Vol 1, pp 1308–1318
66. Wang Q, Liu J, Luo Y, Wang B, Lin C-Y (2016) Knowledge base completion via coupled path ranking. In: Proceedings of the 54th annual meeting of the association for computational linguistics, Vol 1, pp 1308–1318
67. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O (2013) Translating embeddings for modeling multi-relational data. *Adv Neural Inform Process Syst*
68. Nickel M, Tresp V, Krieger HP (2011) A three-way model for collective learning on multi-relational data. In: *Icml*
69. Yang B, Yih Wt, He X, Gao J, Deng L (2014) Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*
70. Liu H, Wu Y, Yang Y (2017) Analogical inference for multi-relational embeddings. In: International conference on machine learning, pp 2168–2178
71. Vashishth S, Sanyal S, Nitin V, Agrawal N, Talukdar P (2020) Interact: improving convolution-based knowledge graph embeddings by increasing feature interactions. *Proc AAAI Conf Artif Intell* 34:3009–3016
72. Dettmers T, Minervini P, Stenetorp P, Riedel S (2018) Convolutional 2d knowledge graph embeddings. In: Proceedings of the AAAI conference on artificial intelligence, vol 32
73. Socher R, Chen D, Manning CD, Ng A (2013) Reasoning with neural tensor networks for knowledge base completion. *Adv Neural Inform Process Syst*
74. Han X, Sun L (2016) Context-sensitive inference rule discovery: a graph-based method. In: Proceedings of COLING 2016, the 26th International conference on computational linguistics: technical papers, pp 2902–2911
75. Toutanova K, Chen D, Pantel P, Poon H, Choudhury P, Gamon M (2015) Representing text for joint embedding of text and knowledge bases. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 1499–1509
76. Zhu Z (2023) A survey of knowledge enabled new generation information systems. *J Softw* 34(10):0–0
77. Singh M, Fuenmayor E, Hinchy EP, Qiao Y, Murray N, Devine D (2021) Digital twin: origin to future. *Appl Syst Innov* 4(2):36
78. Hubauer LS Thomas, Haase P, Herzig DM (2018) Use cases of the industrial knowledge graph at siemens. In: ISWC (P &D/ Industry/BlueSky)
79. Zuheros C, Tabik S, Valdivia A, Martínez-Cámara E, Herrera F (2019) Deep recurrent neural network for geographical entities disambiguation on social media data. *Knowl Based Syst* 173:117–127
80. Ganea OE, Hofmann T (2017) Deep joint entity disambiguation with local neural attention. *arXiv preprint arXiv:1704.04920*

81. Yang J, Yao W, Zhang W (2021) Keyword search on large graphs: a survey. *Data Sci Eng* 6(2):142–162
82. Bornea MA, Dolby J, Kementsietsidis A, Srinivas K, Dantresangle P, Udrea O, Bhattacharjee B (2013) Building an efficient rdf store over a relational database. In: *Proceedings of the 2013 ACM SIGMOD international conference on management of data*, pp 121–132
83. Sun W, Fokoue A, Srinivas K, Kementsietsidis A, Hu G, Xie G (2015) Sqlgraph: an efficient relational-based property graph store. In: *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pp 1887–1901
84. Tian Y, Xu EL, Zhao W, Pirahesh MH, Tong SJ, Sun W, Kolanko T, Apu MSH, Peng H (2020) Ibm db2 graph: supporting synergistic and retrofittable graph queries inside ibm db2. In: *Proceedings of the 2020 ACM SIGMOD international conference on management of data*, pp 345–359
85. Nguyen DQ, Nguyen TD, Nguyen DQ, Phung D (2017) A novel embedding model for knowledge base completion based on convolutional neural network. *arXiv preprint arXiv:1712.02121*
86. Socher R, Huval B, Manning CD, Ng AY (2012) Semantic compositionality through recursive matrix-vector spaces. In: *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pp 1201–1211
87. Takanobu R, Zhang T, Liu J, Huang M (2019) A hierarchical framework for relation extraction with reinforcement learning. *Proc AAAI Conf Artif Intell* 33:7072–7079
88. Naphade M, Smith JR, Tesic J, Chang S-F, Hsu W, Kennedy L, Hauptmann A, Curtis J (2006) Large-scale concept ontology for multimedia. *IEEE Multimedia* 13(3):86–91
89. Arndt R, Troncy R, Staab S, Hardman L, Vacura M (2007) Comm: designing a well-founded multimedia ontology for the web. In: *The semantic web*, pp 30–43
90. Iannacone M, Bohn S, Nakamura G, Gerth J, Huffer K, Bridges R, Ferragut E, Goodall J (2015) Developing an ontology for cyber security knowledge graphs. In: *Proceedings of the 10th annual cyber and information security research conference*, pp 1–4
91. Rastogi N, Dutta S, Zaki MJ, Gittens A, Aggarwal C (2020) Malont: an ontology for malware threat intelligence. In: *International workshop on deployable machine learning for security defense*, pp 28–44
92. Mell P, Grance T (2002) Use of the common vulnerabilities and exposures (cve) vulnerability naming scheme. Technical report, National Inst Of Standards And Technology Gaithersburg Md Computer Security Div
93. Corporation TM (2011) Common configuration enumeration (cce). <http://cce.mitre.org>
94. Mell P, Scarfone K, Romanosky S (2007) Common vulnerability scoring system. *IEEE Secur Privacy* 4(6):85–89
95. Corporation TM (2011) Common attack pattern enumeration and classification (capec). <http://capec.mitre.org/>
96. Zeng J, Zheng LC, Chen Y, Ji K, Mao J (2021) Watson: Abstracting behaviors from audit logs via aggregation of contextual semantics. In: *Network and distributed system security symposium*
97. Fernandez RC, Abedjan Z, Koko F, Yuan G, Madden S, Stonebraker M (2018) Aurum: A data discovery system. In: *2018 IEEE 34th International conference on data engineering (ICDE)*, pp 1001–1012
98. Zheng Y, Li G, Li Y, Shan C, Cheng R (2017) Truth inference in crowdsourcing: is the problem solved? *Proc VLDB Endow* 10(5):541–552
99. Tao X, Pham T, Zhang J, Yong J, Goh WP, Zhang W, Cai Y (2020) Mining health knowledge graph for health risk prediction. *World Wide Web* 23:2341–2362
100. Krishnan S, Wang J, Wu E, Franklin MJ, Goldberg K (2016) Activeclean: interactive data cleaning for statistical modeling. *Proc VLDB Endow* 9(12):948–959
101. Lu J, Chen Y, Herodotou H, Babu S (2019) Speedup your analytics: automatic parameter tuning for databases and big data systems. In: *Very large data bases*