Domain-specific knowledge graphs: A survey

Bilal Abu-Salih

King Abdullah II School of Information Technology, The University of Jordan, Jordan

ARTICLE INFO

ABSTRACT

Knowledge Graphs (KGs) have made a qualitative leap and effected a real revolution in knowledge representation. This is leveraged by the underlying structure of the KG which underpins a better comprehension, reasoning and interpretation of knowledge for both human and machine. Therefore, KGs continue to be used as the main means of tackling a plethora of real-life problems in various domains. However, there is no consensus in regard to a plausible and inclusive definition of a domain-specific KG. Further, in conjunction with several limitations and deficiencies, various domain-specific KG construction approaches are far from perfect. This survey is the first to offer a comprehensive definition of a domain-specific KG. Also, the paper presents a thorough review of the state-of-the-art approaches drawn from academic works relevant to seven domains of knowledge. An examination of current approaches reveals a range of limitations and deficiencies. At the same time, uncharted territories on the research map are highlighted to tackle extant issues in the literature and point to directions for future research.

1. Introduction

KGs, one of the key trends which are driving the next wave of technologies (Bellomarini et al., 2020), have now become a new form of knowledge representation, and the cornerstone of several applications ranging from generic to specific industrial usage cases (Hubauer et al., 2018). The ever-increasing interest in this technology is due to its underlying abstract structure which effectively facilitates domain conceptualization and data management, and its usage as the main driver of several Artificial Intelligence applications. In particular, the KG depicts an integrated collection of real-world entities which are connected by semantically-interrelated relations. In this respect, data are given formal semantics via data annotation and manipulation in a machine-readable format, thereby reducing ambiguity and deriving meaningful information that is specific to an application's domain. Therefore, the incorporation of KGs has extended the existing data models depicted by domain ontologies and established a new form of data analytics that is able to capture semantically-interconnected, large-scale data sets.

Beyond the generic and open-world KGs such as Google KG (Singhal, 2012), most of the current KGs are domain-specific, with certain underlying ontologies in their design (Li et al., 2020a). Because of the lack of a 'one-size fits all' schema or ontology that can be applied to address real-life problems, efforts to establish, polish, and augment domain-specific KGs are continuing to be made in several domains of

knowledge (Kejriwal, 2019). However, this ongoing interest in domain-specific KGs raises questions about their quality and robustness, and whether adequate evaluation measures have been applied, particularly to those KGs derived from data sources of inconsistent quality. Also, the dynamic nature of domain knowledge is highly correlated to contextual situations, and various facts that describe entities might change over time. Neglecting the dynamic nature of knowledge diminishes the quality and correctness of facts represented by KGs, and could lead to poor decision making that is based merely on such data sources. Therefore, it is important that a comprehensive review be conducted of the current state-of-the-art approaches for domain-specific KG construction so as to highlight such issues and address them with viable solutions.

In this survey, we provide an inclusive definition to domain-specific KGs. Further, we discuss various notable KG construction approaches in seven domains of knowledge. These approaches are reviewed and a summary is provided for each domain showing how the KG in each case has been constructed, the resources used to construct the KG, whether any of the KG embedding techniques have been incorporated, the measures used to evaluate the KG construction approach, and the limitations and shortcomings of each approach. This survey paper is different from other similar researches that tend to either focus on generic and domain-independent KGs such as (Ji et al., 2020; Paulheim, 2017; Wang et al., 2017a), or only briefly discuss domain-specific KGs (Kejriwal, 2019; Zou, 2020). To the best of our knowledge, this is the

first attempt to provide both an inclusive definition of a domain-specific KG as well as a thorough analysis of various domain-based KG construction approaches. Further, this paper provides a summary of the main issues derived from the conducted analysis. At the same time, several improvements, recommendations, and opportunities are suggested to address these limitations which point to directions for future research. The contributions of this paper are:

- To the best of our knowledge, this is the first paper to provide an inclusive definition of a domain-specific KG and the first comprehensive survey of domain-specific KGs.
- We conduct a thorough analysis of more than 140 papers on KG construction approaches, covering seven domains.
- The review reveals the shortcomings of current approaches and proposes solutions to address them.
- The paper highlights research gaps in the area of domain-specific KG construction and suggests venues for future research.

The rest of this paper is organised as follows: Section 2 discusses the methodology adopted in this study. Section 3 establishes the necessary ground for this work by including important preliminaries and relevant terminologies. Section 4 analyses the KG construction approaches in seven domains. Section 5 presents an in-depth discussion of the findings as well as the identified research gaps. Section 6 concludes the paper.

2. Methodology

As the first step, seven domains were chosen and the articles for review were collected accordingly. The domains are: healthcare, education, ICT, science and engineering, finance, society and politics, and travel. Articles relevant to any of these domains were obtained by searching recent volumes of both conference proceedings of relevant series (such as ACM SIGKDD, ACM WSC, WWW, ICWE, ISWC, etc.) and high-quality journals (such as Knowledge-based Systems, Expert Systems with Applications, IEEE Access, Journal of Web Semantic, etc.). Furthermore, we used keywords such as "knowledge graph for engineering", "knowledge graph for healthcare", etc. to search for articles in Google Scholar. Although the articles chosen for each domain intuitively depict the KG construction approaches to the domain of interest, it was noted that some papers were concerned with more than one high-level domain of knowledge. Those papers dealing with interrelated domains are positioned under the domain umbrella that is explicitly indicated by the papers' authors.

We examined more than 140 research articles that appeared between 2016 and 2020 in high-quality computer science and information systems publication venues. Fig. 1 indicates the number of papers collected for the review. As seen in the figure, the interest in domain-specific KGs has increased dramatically over recent years.

Our study is different from other similar works. For example, the current important researches in this arena, such as (Ji et al., 2020; Paulheim, 2017; Wang et al., 2017a), either focus on generic and domain-independent KGs, or only touch upon and do not discuss in depth domain-specific KGs (Kejriwal, 2019; Zou, 2020). To the best of our knowledge, this is the first attempt to provide both an inclusive definition of the term 'domain-specific KG' as well as a thorough analysis of various domain-based KG construction approaches. Furthermore, the review reveals the shortcomings of current approaches and proposes solutions to address them.

3. Preliminaries

3.1. Generic knowledge graphs

Generic KGs (a.k.a. open-world, cross-domain, or domain-independent) were constructed long before the term "knowledge graph" was coined, and such constructions have been ongoing. In fact,
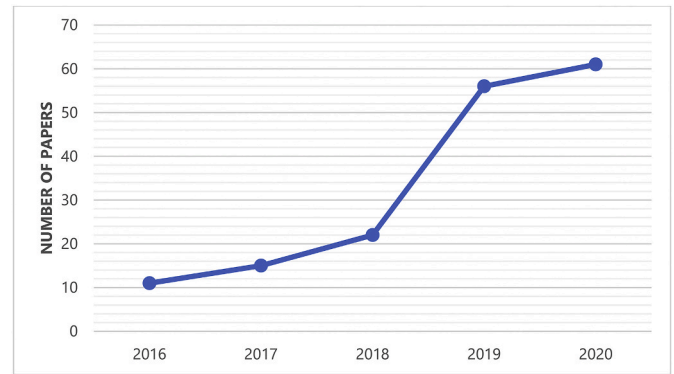

Fig. 1. Number of studied papers per year.

since the invention of the Semantic Web, generic KGs have been associated with linked data, being natural representations of the interlinking of entities (Ehrlinger and Wöß, 2016). Nevertheless, the term has gained much momentum recently as it has given rise to new computing paradigms by shifting from traditional databases to knowledge databases (Menon et al., 2019). Ironically, there is no consensus on the definition of the term despite the few attempts to provide a reasonable description. For example, Ehrlinger and Wöß (2016) perceive the KG as the process of acquiring and correlating knowledge to an ontology and applying a reasoner to infer knowledge. A further technical depiction of the term is provided by Wang et al. (2017a) who conceive the KG as a multidimensional graph comprising entities/nodes and relations/edges.

A KG is commonly described as a directed graph (G), where $G = (V, E)$. This notation depicts the relationship between vertices (V) of the graph and edges (E) between these vertices. The vertices represent the set of real-world entities and the edges represent the relationships between these entities. Vertices/entities/nodes are interconnected using relations which are the edges of the graph, and facts are commonly represented as an RDF[1] triple (*subjects, predicate, object*) or (*head, relation, tail*), and denoted as $< h, r, t >$. Intuitively, two entities connected by a relation form a fact in the KG. For example, Fig. 2 gives an example of a KG representation in terms of entities and relations.

Several facts (triples) can be inferred from this KG. For example, the facts "*Tim Berners-Lee has invented WWW*" comprises two entities/nodes, namely "*Tim Berners-Lee*" and "*WWW*", and the relation "*has invented*" forms the triple "*Tim Berners-Lee, hasInvented, WWW*".

Examples of various steadily evolving open-world KGs include: BabelNet,[2] YAGO,[3] Cyc,[4] NELL,[5] CliGraph[6] and DBPedia[7] knowledge base. In fact, several of these massive publicly-available data islands have been harvested from the Web as key sources of knowledge to benefit numerous Artificial Intelligence and smart systems (Wang et al., 2019a), such as recommender systems (Catherine and Cohen, 2016), decision support systems (DSSs) (Elnagar and Weistroffer, 2019), and intelligent QA systems (Yang et al., 2019a). Table 1 presents statistics for popular generic KGs. The statistics include #instances, #facts, #types (indicating the number of classes in the designated schema), and #relations.

3.2. Domain-specific knowledge graphs

Despite the extensive use of the generic and open-world KGs to tackle

---

[1] https://www.w3.org/TR/rdf11-concepts/.
[2] https://babelnet.org/.
[3] http://www.foaf-project.org/.
[4] https://www.cyc.com/.
[5] http://rtw.ml.cmu.edu/rtw/kbbrowser/.
[6] http://caligraph.org/ontology/Scientist.
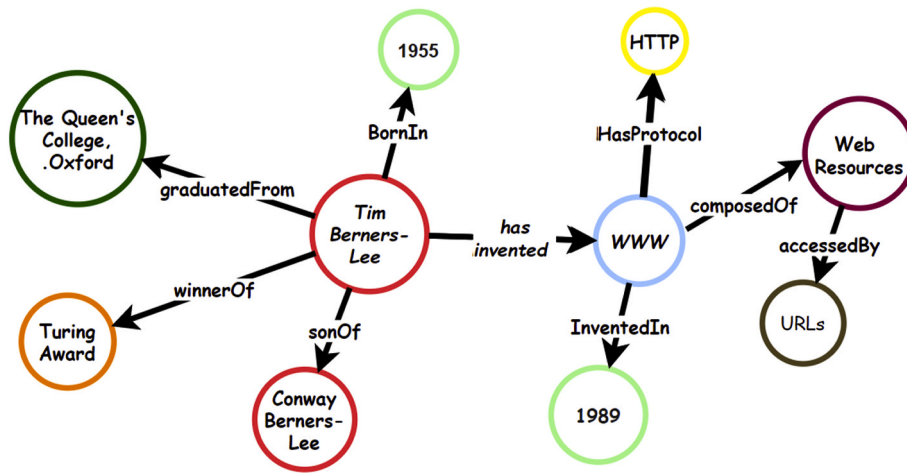[7] https://wiki.dbpedia.org/.

Fig. 2. An example of entities and relations in a KG.

Table 1
Statistics for some of the open world KGs (Heist et al., 2020).

	Instances/entities	Facts	Types/classes	Relations
DBpedia	5,044,223	400	760	1355
YAGO	6,349,359	479,392,870	819,292	77
Wikidata	52,252,549	732,420,508	2,356,259	6236
BabelNet	7,735,436	178,982,397	6,044,564	22
Cyc	122,441	2,229,266	116,821	148
NELL	5,120,688	60,594,443	1187	440
CaLiGraph	7,315,918	517,099,124	755,963	271
Voldemort	55,861	693,428	621	294

a wide variety of domain-independent tasks, constructing KGs from domain corpora to address domain-specific problems is greatly important (Kejriwal et al., 2019). This is because domain-specific KGs have relevant and semantically interlinked applications with domain-specific problems. Moreover, domain-specific KGs also lack a well-established, consensual and comprehensive definition, which is not surprising given that this is still a comparatively new territory and an under-explored frontier (Shen et al., 2019). Nevertheless, some studies perceive the domain-specific KG as a special type of KG that is used to represent a specific and complex domain (Li et al., 2020a; Yuan et al., 2020; Fan et al., 2017). Others reported that domain-specific KGs are the result of the process of enriching an underlying domain ontology (Kejriwal, 2019). This inability to provide an inclusive definition to the domain-specific KG has driven us to frame the following definition of this term:

> "Domain Knowledge Graph is an explicit conceptualisation to a high-level subject-matter domain and its specific subdomains represented in terms of semantically interrelated entities and relations".

This comprehensive definition addresses three core aspects, namely: (i) formal conceptualization: which indicates the logical design of the KG depicted by a specific and predefined domain ontology established to capture the domain of interest either in its generic (high-level) sense or in its specific subdomains; (ii) subject-matter domain: this frames the domain-specific KG to be firmly contextualised to address a particular subject-matter knowledge; and (iii) semantically interrelated entities and relations, which indicate the physical design of the domain-specific KG depicted as a labelled graph in which the semantics of data is enriched with a specific conceptual representation of entities and relationships between these entities.

3.3. Knowledge graph construction

KG was introduced as an efficient and smart approach to tackle the continuous propagation of various forms of unstructured text (e.g. Web data) and other structured or semi-structured sources (Ji et al., 2020). The construction of a KG can be perceived as a paradigm which comprises different perspectives. We have examined the literature on KG construction approaches and we design a taxonomy to highlight the key aspects of KG construction. Fig. 3 illustrates the taxonomy for KG construction which is mainly categorised based on the level of knowledge extraction, the type of knowledge base, and the incorporated construction method.

In terms of knowledge extraction, KG construction is the process of extracting entities and relationships between these entities. Entity extraction comprises three key tasks (Al-Moslmi et al., 2020), namely: i) Named Entity Recognition (NER) which involves the process of finding individuals, organisation, locations, events, and other entities from (un) (semi-)structured data sources; ii) Named Entity Disambiguation (NED) which aims to eliminate the ambiguity of an inferred entity by mapping it to the factual real-world entity; and iii) Named Entity Linking (NEL) which assigns a unique IRI identified to the disambiguated entity. The purpose of relation extraction is to discover the semantic relationships between the identified and disambiguated entities. Relation extraction can be done by either local-based or global-based relation extraction methods. The former indicates the mention-level relation which is commonly inferred form short textual contents, and the latter aims to infer those relations that span several local-based relations. The reader can obtain further details on these tasks in relevant literature such as (Kim et al., 2020; Smirnova and Cudré-Mauroux, 2018).

Another perspective conceives that the KG construction depends on whether a predefined ontology schema is used (schema-based), or the KG is constructed with no predefined schema (schema-free) (Nickel et al., 2015), or if the construction comprises a hybrid of schema-based and schema-free approaches. The first methodology (schema-based) can be categorised according to two groups based on the selection of data sources and ontology (Dong et al., 2014; Heist, 2018) namely: (i) the bottom-up method whereby the structural nature of an ontology is incorporated to build the KG (e.g. Wikipedia is established by using the predefined ontology model, i.e. DBpedia (Kuhn et al., 2016)); and (ii) the top-down method where the ontology schema is inferred from the underlying structured data (e.g. YAGO) (Suchanek et al., 2007)) or build taxonomies (hierarchy) are constructed based on information from the Web (Wu et al., 2012). Schema-free approaches are open information extraction techniques which rely on the openness of the Web; hence, the information is collected using various knowledge extraction techniques with no consideration given to creating a unified ontology design (e.g.
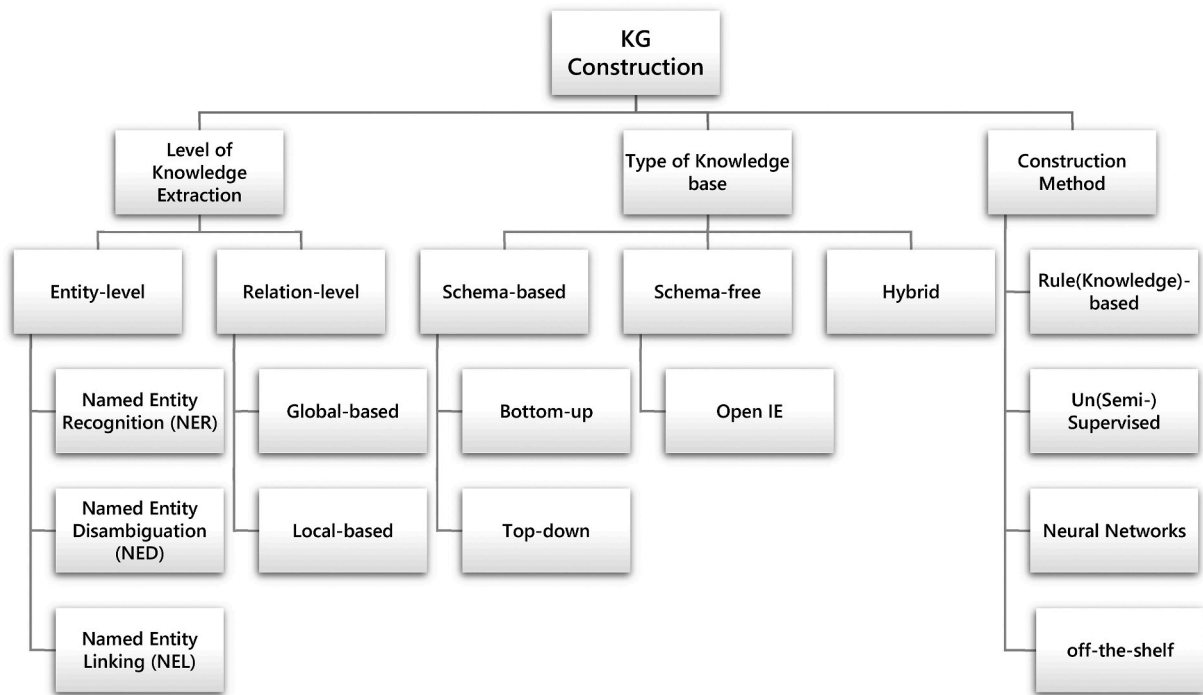
Fig. 3. A taxonomy for KG construction.

OpenIE (Fader et al., 2011)). Hybrid knowledge-based approaches: are the techniques whereby the knowledge is obtained based on a pre-determined ontology, as well as facts gathered from the Web (e.g. KnowledgeVault (Dong et al., 2014), NELL (Carlson et al., 2010)).

The third thread of research sees the KG construction as being based on the technical solutions and methods used to mine the Web and other data silos and repositories to deduce entities and relations, thereby constructing the KG. Various approaches were proposed in the literature to address this issue. These fall under four key categories (Ji et al., 2020; Al-Moslmi et al., 2020; Smirnova and Cudré-Mauroux, 2018): (1) Knowledge-based approaches: these methods mainly involve domain experts to establish human-crafted rules (Association Rules) by incorporating domain-knowledge and lexicon resources. (2) Learning-based methods: these methods encompass (un)(semi-)supervised techniques. Examples of these methods that are employed for entities recognition and relations extraction are Conditional Random Field (CRF), Hidden Markov Models (HMM), Support Victor Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR), and Decision Trees, and bootstrapped methods. (3) Neural network models: these methods incorporate neural networks and deep learning to infer features. These methods require neither domain ontologies nor domain-specific corpora, thereby making them suitable for domain-independent KGs. Examples of these methods Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Bidirectional Long Short-Term Memory (BiLSTM). (4) Off-the-shelf NLP tools: those are commonly commercial/open source ready-to-use tools which might embed one or more of the aforementioned approaches to analyse textual content and extract entities and relations. Examples of these tools are spaCy,[8] Stanford CoreNLP,[9] AllenNLP [10], and IBM Watson NLU.[11]

3.4. Knowledge graph embedding

KG Embedding (KGE) is the process of creating propositional feature vector representations of the constituents of a KG (entities and relationships) (Wang et al., 2017a) so as to apply numeric techniques that produce scalable and effective results (Bonatti et al., 2019). KGE techniques can simplify the resolution of various complex real-life graph problems for which conventional graph presentation (i.e., adjacency matrix) is inadequate and inferior. Hence, currently, KGE is being used extensively to tackle problems such as KG completion, entity recognition, and link-based clustering (Nickel et al., 2015; Kipf and Welling, 2016; Wang et al., 2017b).

The core idea underlying KGE is to create a vector for each entity and each relation in the KG, then define a set of score functions that are used to measure the space distance of two entities relative to the type of the relation in the low-dimensional embedding vector space. The aim is to capture latent properties of the semantics in the KG so that like entities and like relations will be represented with similar vectors, and those not semantically connected are detached. The embedding of a KG is learned via training a neural architecture over a KG, and usually includes three main steps (Bianchi et al., 2020), namely: (i) encoding entities into dispersed points in the semantic space, and encoding relations as vectors; (ii) scoring function or model-specific function that is used to assemble the information combing from a triple; and (iii) applying an optimization procedure represented by the loss function in which the objective is defined and minimised during the KG embedding process. Table 2 shows the most popular of the many models which are commonly used for KG embeddings.

3.5. Knowledge graph evaluation

The proliferation of massive KGs poses a question regarding the quality of the embedded knowledge (i.e., entities and relations), and whether these facts precisely convey the intended real-world concepts interlinked via their relationships. Therefore, ascertaining the completeness and correctness of the constructed KG is crucial to determine its "fitness of purposes" (Chen et al., 2019a) for various

---

[8] https://spacy.io/.
[9] http://corenlp.run/.
[10] https://allennlp.org/.
[11] https://www.ibm.com/cloud/watson-natural-language-understanding.

Table 2
Examples of well-known KG embedding models.

Model	Description	Scoring Function
TransE (Bordes et al., 2013)	learns the representation of both the entities and relations as vectors in the same low dimensional semantic space. Hence, for a golden triple (h,r,t), **TransE** treats the relation r as a translation in the embedding space so that $h + r \approx t$, when (h,r,t) holds (t should be the closest to $h + r$), otherwise $h + r$ should be away in distance from t.	$f_{TransE} = -\|e_h + e_r - e_t\|_n$
DistMult (Yang et al., 2014)	is an extension and a simplification to RESCAL (Nickel et al., 2011) and is based on the bilinear model. In this model, the relation is encoded as diagonal (single vector) using the trilinear dot product.	$f_{DistMult} = r^T(h \odot t)$
ComplEx (Trouillon et al., 2016)	this is an extension to DistMult model by introducing complex-valued embeddings, where the scoring function is based on the trilinear Hermitian dot product in C.	$f_{ComplEx} = Re(e_r, e_h, \overline{e_t})$
HolE (Nickel et al., 2016)	a compositional vector space model that learns compositional vector space representations of entities and relations through incorporating the strength of RESCAL as well as the simplicity of **DistMult.**	$f_{HolE} = w_r \cdot (e_h \otimes e_t) = \frac{1}{k}F(w_r) \cdot (\overline{F(e_h)} \odot F(e_t))$
ConvE (Dettmers et al., 2018)	is a neural link prediction model that uses deep, multi-layer, conventional and fully connected layers of nonlinear features to tackle the interactions between input entities and relations.	$f_{ConvE} = \langle \sigma(vec(g([\overline{e_h}; \overline{e_r}] * \Omega))W))e_t \rangle$
ConvKB (Nguyen et al., 2017)	incorporates conventional neural networks to represent the concatenation of entities and relations, which increases the learning ability of latent features.	$f_{ConvKB} = concat(g([e_h, e_r, e_t] * \Omega)) \cdot W$

downstream applications, and to deal with the uncertainty in the data quality (Gao et al.).

In a domain-specific KG, the absence of a complete and accurate KG poses a challenge to the evaluation process. This is because collecting all true facts about a certain domain of interest is not a trivial task (if not impossible). Therefore, various attempts, commonly known as KG Augmentation/Completion techniques, have been undertaken to augment the KG with new facts presented by new potential entities and/ or new relations. To ensure data quality, these efforts are subject to correctness and completeness evaluation measures. In particular, according to the new and known true facts, the evaluation can be carried out by using classification accuracy and ranking metrics such as Hits@N and Mean Reciprocal Rank (MRR), Accuracy, Precision, Recall, and F-score (Kejriwal, 2019; Pezeshkpour et al., 2020). These metrics are amongst various other measures that are currently being incorporated to evaluate the KG construction and completion in terms of the factuality of the embedded entities as well as their relations. Table 3 lists some of the well-known evaluation metrics.

Evaluations of KG constructions have also been carried out by means of case studies and domain experts (Zaveri et al., 2016; Mohammadhassanzadeh et al., 2018).

4. Domain-specific KGs

This section reviews various domain-based KGs that were discussed in the literature. These domains are Healthcare, Education, ICT, Science and Engineering, Finance, Society and Politics, and Travel. Appendix A includes seven tables, each with a summary of the literature for each designated domain. These tables demonstrate the specific KG usage, KG construction algorithm(s), the resources used to feed the KG, whether KG embedding techniques were incorporated, the evaluation approach, and the limitations of each designated work.

4.1. Healthcare

Recently, the Healthcare sector has gained much attention, particularly with coronavirus 2019 (COVID-19) pandemic continues to rattle the world. Therefore, there is a notable consensus in both industry and academia to consolidate efforts to overcome the challenges of this vital sector (Cui et al., 2020). KGs offer the healthcare sector technical means to derive meaningful insights from voluminous and heterogeneous healthcare data (Zhang et al., 2020a; Malik et al., 2020). The examined papers in Healthcare domain can be categorised into the following subdomains.

4.1.1. Generic healthcare

Health data mining by means of a KG approach was followed in the literature. Gatta et al. (2017) presented a library in R that was developed for process mining in the medical domain. The library is designed to encode the extracted processes in the form of directed graphs, which can be then interpreted and visualised by domain experts. Another important effort that integrated plausible reasoning with fine-grained

Table 3
Examples of well-known KG Evaluation metrics.

Metric	Description	Formal definition	
Hits@N	indicates the number of elements in the ranking vector retrieved from the model is positioned in the top (N) locations.	$Hits@N \begin{cases} \sum_{i=1}^{\|Q\|} 1, if\ rank(s,p,o)_i \leq N \\ 0, \quad otherwise \end{cases}$	where $rank(s,p,o)_i$ refers to the rank of a positive element i against a list of negative elements, T is a set of test triples and (s,p,o) is a triple $\in T$.
MRR	is a function that computes the mean of the reciprocal of elements embodied in a vector of rankings. It is used as a measure to evaluate the system performance against the retrieved elements.	$MRR = \frac{1}{\|T\|}\sum_{i=1}^{\|Q\|} \frac{1}{rank(s,p,o)_i},$	
MR	refers to the mean rank of the correct test facts/triples embodied in a vector of rankings (i.e., the average of the predicted ranks).	$MR = \frac{1}{\|T\|}\sum_{i=1}^{\|Q\|} rank(s,p,o)_i$	
Accuracy	specifies the accuracy of the incorporated embedding model in making a correct prediction. Accuracy is the ratio obtained between the accurate predictions (i.e., $TP + TN$) and the overall inferred predictions ($FN + TP + FP + TN$).	$Accuracy = \frac{TP + TN}{FN + TP + FP + TN}$	True-Positives (TP): refer to the true facts that are classified by the model as true statements. False-Positives (FP): refer to the false facts that are classified incorrectly as true statements. True-Negatives (TN): refer to the false facts that are classified correctly as false statements. False-Negatives (FN): refer to the true facts that are classified incorrectly as false statements.
Precision	refers to the proportion of those facts that were classified accurately as positive and they are actually positive.	$Precision = \frac{TP}{TP + FP}$	
Recall	recall indicates the proportion of true positive facts were correctly classified as positives facts.	$Recall = \frac{TP}{TP + FN}$	
F- measure	(a.k.a. f-measure/f-score) is a harmonic measure used to provide a trade-off between precision and recall	$F - score = \frac{Precision \cdot Recall}{Precision + Recall}$	

biomedical ontologies to tackle data incompleteness problem in generic health data was undertaken by Mohammadhassanzadeh et al. (2018). The authors proposed Semantics-based Data analytics (SeDan) framework that performs an exploratory and plausible analysis of the KG using plausible OWL extension and query rewriting algorithm. The framework incorporates various knowledge bases including the DrugBank, Disease Ontology, and the large-scale semantic MEDLINE database (SemMedDB). Rastogi et al. (Rastogi and Zaki, 2020) framed the personal health KG as a combination of context, personalization, and integration with other knowledge-bases. Their study indicated that the literature on personalized health-related KGs is inadequate and lacks a unified standard representation to depict the designated domain. Incorporating health KGs for Query Answering (QA) system was discussed by Sahu et al. (2018). The authors offered a system that can be used to search for various health-based KGs to obtain a set of healthcare-related response sub-graphs. Incorporating KG in the medical domain to benefit QA applications was also discussed in Sheng et al. (2020).

4.1.2. Diseases

The advances of constructing KGs to conceptualise health-related problems and diseases were discussed in the literature. Rotmensch et al. (2017) constructed a KG that captures diseases and symptoms related entities form 273,174 electronic medical records. The authors incorporate Google Health Knowledge Graph (GHKG) and created a KG that embodies diseases and symptoms and relationships between them. Constructing KGs that can describe depression was provided by Huang et al. (2017a). In particular, they attempted to generate a sub-graph, that describes depression disorder, obtained by parsing a variety of large knowledge sources such as PubMed, Medical Guidelines, DrugBank, Unified Medical Language System (UMLS) etc. Evaluating the robustness of a constructed KG in healthcare is of utmost significance to ensure the quality of the inferred knowledge in this sensitive domain. In this context, Chen et al. (2020a) presented a methodology to measure and evaluate the robustness of knowledge in terms of diseases and symptoms captured from existing health knowledge graphs as well as records of patient visits to the Beth Israel Deaconess Medical Center (BIDMC). Addressing the temporal dimension in KG creation is an important dimension in healthcare. Ma et al. (2019) established a temporal KG that can be used for cognitive episodic memory. This temporal KG was initially derived from the Integrated Conflict Early Warning System (ICEWS) dataset as well as Global Database of Events, Language and Tone (GDELT). Their work was different from other seminal works by generalizing four significant static KGs embedding to 4-dimensional temporal/episodic KGs. Also, two novel generalizations of RESCAL were proposed and discussed. Application of KGs in healthcare and medical domains was demonstrated in other relevant tasks such as fraud, waste, and abuse Detection (Sun et al., 2020), drugs similarity (Shen et al., 2019), drug repurposing (Zhu et al., 2020), clinical decision support systems (Li et al., 2020b), and medical recommender systems (Sadeghi and Lehmann, 2019; Lin et al., 2020).

4.1.3. Healthcare management

Constructing a KG to benefit health management and to address current health-related problems and chronic diseases were proposed in the literature (Huang et al., 2019; Haussmann et al., 2019; Chi et al., 2018a; Zhang et al., 2019a). For example, Huang et al. (2019) suggested a KG construction model that benefits people seeking knowledge regarding a healthy diet. The authors proposed a domain ontology as an underlying structure of a diet KG. The KG was then enriched with entities extracted from a set of healthcare websites using Conditional Random Fields (CRF), Support Vector Machine (SVM) and Decision Tree (DT) algorithms. Another effort was carried out by Haussmann et al. (2019) who proposed an integrated KG (FoodKG) that embodies knowledge on healthy food, recipes, and nutrition. The authors ensured the credibility of the obtained knowledge by adopting RDF Nano-publication specification (Groth et al., 2010). On the same research

direction, an inclusive healthy diet KG was also constructed by Chi et al. (2018a). The reported KG integrated five key concepts that included food material, dish, nutritional element, symptom, and the crowd. Through semi-automatically extraction approach, the proposed model was capable to collect and import entities captured from a set of online resources using various NLP and machine learning algorithms. Modeling food domain-specific KGs were also implemented in Yu et al. (2020); Zeng and Nakano (2020); Salehian et al. (2019). Further, tackling challenges in healthcare systems leveraging KGs technologies was discussed in Wang et al. (2020a); Tao et al. (2020a); Goodwin and Harabagiu (2017).

4.2. Education

The construction and usage of educational KGs have been extended recently due to the significance of KGs application to the learning systems as well as the abundance of pedagogical data (Shi et al., 2020). The following subdomains are inferred from the collected education-related papers.

4.2.1. Teaching and classroom resources

KGs have proven ability to foster learning (Cui and Yu, 2019) and been used in popular massive open online course (MOOC) platforms (Chen et al., 2018a; Zheng et al., 2017). For example, Chen et al. (2018b) presented K12EduKG, a KG constructed based on K-12 educational subjects. Domain-specific educational data (*Chinese curriculum standards of mathematics*) was the source of knowledge that was used in K12EduKG. Concepts and relations are identified and imported into K12EduKG using CRF model and probabilistic association rule mining. Su and Zhang (2020) designed a KG schema that can accommodate educational Big data. Their KG was enriched by using two large datasets, namely subject teaching resources and an online encyclopedia resource. Another attempt has been undertaken to build a KG (MathGraph) that can be used to solve high school mathematical exercises including mathematical derivation and calculation (Zhao et al., 2019). The authors of Zhao et al. (2019) used MathGraph, that was initially constructed with the help of crowdsource, to embody dissimilar mathematical objects, operations and constraints. KnowEdu (Chen et al., 2018a) is one of the important efforts in educational KG design and construction. By using standard curriculum and learning assessment data as data sources, and by using neural network models for concepts and relations identification, KnowEdu is created to facilitate learner's cognitive and educational process.

4.2.2. Education management

Aliyu et al. (2020) presented an approach for implementing a KG to be used for course allocation scheduling and management. Although the evaluation of the constructed KG was inadequate, the work is promising toward this important research direction. Liu et al. (2016) and Lian et al. (Liang et al., 2018) provided approaches that adopted graph-based relational learning for concept prerequisite learning in education domain. The former introduced an automatic technique for prerequisites prediction by inferring directed graph at the course and concept levels. The latter incorporated active learning to the concept prerequisite learning problem. Modelling internal control in higher education using KG technology was discussed in Wang (2020). The author proposed a KG that can conceptualise the internal control policy in the higher educational institutions. Despite the author's attempt to demonstrate the utility in a visualisation task, the overall mechanism to construct and evaluate the designated KG is inadequate and inferior. Applying KGs to benefit education management was also outlined in Zhang and Zhang (2020).

4.2.3. Educational technologies

Designing a KG that can be used to depict academic networks was discussed in Chi et al. (2018b). The authors proposed a model of

scientific publication management that can integrate scientific metadata in terms of academic entities. The model embodies a KG which conveys the relationship between research entities and research topics. Incorporating KG embedding techniques in the process of constructing the educational KGs is inadequate, particularly those leverage the rich literals of the designated KGs, thus Yao et al. (2020) attempted to tackle this issue by reporting on a model for embedding learning of educational KGs. With the use of three experimental KGs in the education domain, the authors demonstrated the significance of the proposed model when processing educational KGs. In particular, authors of Yao et al. (2020) presented a method that can jointly learn embeddings built on pre-trained structural (i.e. TransE) and literal embedding vectors (i.e. BERT). Evaluating the utility of the educational KG by means of visualisation analysis was rarely discussed in the literature. An attempt in this direction was undertaken by Sun et al. (2016) who integrated an education KG and demonstrated its utility by carrying out visual analysis so as to provide a better understanding to its topological structure.

4.3. ICT

KGs have been widely used to improve several subdomains related to Information and Communication Technology as follows:

4.3.1. Cybersecurity

Detecting and preventing cyberattack is inevitable to ensure providing continuous and uninterrupted services. Interestingly, various cybersecurity-related KGs have been introduced and developed. For example (Jia et al., 2018), presented a practical approach for cybersecurity. They first developed a domain ontology that put forward a technique to construct the cybersecurity KG. Then they proposed a quintuple model that was used to obtain new knowledge using the path-ranking algorithm. Deng et al. (2019a) discussed another cybersecurity-related KG that was constructed to serve students who seek concepts in this domain. Despite the Adhoc mechanism to construct their KG and the absence of a benchmark comparison for utility evaluation, the line of research is important per se; providing KGs that facilitate personalized learning and benefit education is highly recommended (Meissner and Köbis, 2020). Kiesling et al. (2019) followed a bottom-up approach to build their cybersecurity KG using National U.S. Vulnerability Database (NVD) and set of security online references. The authors demonstrated the effectiveness of the developed KG by means of two case studies in vulnerability assessment and intrusion detection systems. Cybersecurity KGs were also reported in Pingle et al. (2019); Tao et al. (2020b); Hooi et al. (2019).

4.3.2. Software development

Software development is a sophisticated process that encompasses an array of challenges and decisions to be made in a timely manner (Abu-Salih et al., 2021a). Hence, the software engineering domain has also benefited from the propagational use of domain-specific KGs due to their efficacy to store and manage relevant entities and relations of high complexity. For example, Nayak et al. (2020) developed a KG that was used to extract test cases that would assist in the functional requirements gathering process. As a backbone schema, the authors designed an ontology for software testing and applied a series of NLP tools including Constituency Parse Tree (CPT) to mine and populate the KG with test cases. Schindler et al. (2020) introduced a KG (SoftwareKG) that embodies information pertaining to the software mentioned in academic articles of social science. SotwareKG depicts various aspects of the software including its availability, source and links with other knowledge repositories. Designing a framework to industrial software design and development processes was proposed by Li et al. (2019). The authors applied a knowledge-driven QA system for parameters searching and can be also used for carrying out variable calculation and ontology reasoning. The proposed model integrated the constructed KG with an SQL database and efficiency is demonstrated in certain industrial

scenarios. Fu et al. (2019) made use of IT crowdsourcing services to construct a KG (ITServiceKG) to improve existing IT services IR system. The authors implemented a learning-to-rank model (Gradient Decision Tree) that was leveraged to re-rank the obtained results, thereby attaining much relative search results. Constructing KGs to improve software engineering practices and internal processes were also addressed in Wang et al. (2019b); Xiao et al. (2019); Zhang et al. (2020b); Xie et al. (2019).

4.3.3. Telecommunication

Defining and conceptualizing the structure of telecommunication networks can be explicitly delivered through incorporating KGs. Aumayr et al. (2019) benefited from the unique structure of KG to build a graph that can be used to solving issues that confront telecom operators. The authors populated the KG with entities captured form community knowledge in forms of telecom and products documentations, online sites, engineering reports, etc. The aim was to build an automated system to improve network incident management processes and to provide better customer service. Krinkin et al. (2020) proposed a telecommunication network monitoring model by means of incorporating a domain-specific KG at the top of the telecommunications service domain ontology. In particular, the authors encompass an array of components that are integrated into monitoring cable television operator networks. These components are: the billing model, user access rights, network topology and application hierarchy, and the cable television operator network service model.

4.3.4. Internet of things (IoT)

In IoT integrating heterogeneous access of electronic devices poses a momentous challenge. Hence, the underlying structure of the KG offers a promising solution to bridge the gap between IoT devices. For example, Xie et al. (2020) proposed an IoT KG that was used in a new layer to map IoT devices, thereby unifying the communications of all devices. Many studies further elaborated on KGs and their advantages to the IoT ecosystem (Le-Phuoc et al., 2016; Zhang et al., 2016; Chen et al., 2019b).

4.4. Sciences and engineering

Applying semantic web technologies and ontologies in natural sciences has proven successful leveraging the formal knowledge representation and the semantic web languages that can model rich and complex knowledge of natural sciences (Hastings et al., 2012; Taylor et al., 2006; Collins and Clark, 2014; Dupre, 2014; Raskin and Pan, 2005). Examples of subdomains that benefited from the KG technology are reviewed as follows:

4.4.1. Chemistry

The utility of semantic analytics has been validated in providing a formal representation to chemical data, thereby increasing the sharing and interoperability of such data (Farazi et al., 2020a). Incorporating KG has extended these endeavours and provided a platform where information can be integrated from multiple chemical kinetic systems, and offered an automatic method to comprehend chemical mechanisms to perform complex chemical-related semantic queries (Farazi et al., 2020b). For example (Farazi et al., 2020c), proposed an integrated system that used KG to demonstrate interoperability in cross-domain applications that compass combustion as well as to address the problem of data inconsistencies in chemical reaction mechanisms. Krdzavac et al. (2019) designed a domain ontology (OntoCompChem) as an underlying structure of a KG that was used to demonstrate quantum chemistry calculations.

4.4.2. Biology

The continuous propagation of the throughput data in molecular biology has necessitated the introduction of formal representation of this knowledge domain. For example, Humayun et al. (2020) developed a

mechanistic KG to depict the heme's interactome, an important factor of diverse biological processes. Choi et al. (Choi and Lee, 2019) carried out benchmark comparison amongst a selection of KG embedding models in a relational discovery task. Prior to the implementation of the incorporated KG embeddings, the authors constructed a domain-specific KG that was populated with entities and relations captured from heterogeneous public data sources. These bio-related data sources are PubMed database,[12] Comparative Toxicogenomics database (CTD),[13] The Biological General Repository For Interaction datasets(BioGRID),[14] and the human disease database: (MalaCards).[15]

4.4.3. Geology

Incorporating KGs on geological data has proven effectiveness and enhanced the interconnectivity between such data sets. Zhu et al. (2017) showcased the use of KG in an intelligent system for deep mining of geological data. The authors constructed the KG using Baike.com and local geological documents. Constructing a KG from geoscience documents was discussed in Wang et al. (2018a). The authors made use of an integrated corpus composed of a geology dictionary and the *Terminologies and Classification Codes of Geology and Mineral Resources (*TCCGMR) (Xinqing et al., 1999) to enrich the KG incorporating CRF-base geological word segmentation model. The utilization of KG to conceptualise geosciences were further detailed in (ZHANG et al., 2020; Bucher et al., 2020; Fan et al., 2020a).

4.4.4. Engineering

More advanced technical solutions have been also importing KGs to build sophisticated systems in various fields of engineering (Fan et al., 2019). For example, Myklebust et al. (2019) demonstrated the implication of using a domain-specific KG and KG embeddings to improve the ecotoxicological effect prediction in the Norwegian Institute for Water Research (NIVA). In particular, the authors designed TERA KG to integrate information captured from dissimilar resources relevant to ecotoxicology and risk assessment domain (e.g. ECOTOXicology database (ECOTOX)[16] and NCBI taxonomy[17]). The construction of TERA KG was carried out by using LogMap ontology alignment system (Jiménez-Ruiz and Grau, 2011) to index and align the ECOTOX and NCBI vocabularies. Yan et al. (2020) designed KnowIME (KG's Intelligent Manufacturing Equipment), a knowledge-based integration system for manufacturing equipment such as lathes, conveyors and robots. A domain-specific KG was constructed and augmented using CRF method from heterogeneous equipment-related data. Industrial adoption to KGs to enhance customer satisfaction and user experience was undertaken by Li et al. (2020c). The authors designed two KGs for evolutionary Smart Product– Service System (Smart PSS) development. The constructed KGs resulted from data obtained from open-source knowledge, prototype specifications, and user-generated textual data. The KGs were then used to bringing expert knowledge, thereby solving issues related to cost-effectiveness that exist in the current knowledge supply for Smart PSS.

Electric power artificial intelligence systems have also contributed to the KGs construction and augmentation. For example, Fan et al. (2020b) proposed an approach to construct the dispatch KG for the power grid to semantically describing behavior of dispatchers. They follow semi-automated labelling to construct a power corpus, then BiLSTM-CRF model was used to extract entities and indicate the dispatching behavior relationship patterns. Yang et al. (2019b) also leveraged the KG schema to collect and integrate data from various power assets. They aimed to deliver a unified multi-source

heterogeneous knowledge base that contains power transmission and transformation assets. Engineering is a wide spectrum of domains that have found KGs advantageous in several important applications. This is evident in various other engineering fields such as, nuclear engineering (Zhao and Smidts, 2019), marine engineering (Fujun et al., 2017), Photonics engineering (Li et al., 2020d), Nanotechnology Engineering (Erkimbaev et al., 2019), Ceramics engineering (Gao et al., 2019), and Geomatics engineering (Jiang et al., 2019).

4.5. Finance

The finance sector is a pillar of any successful business. It is the driver for businesses to take opportunities and make revenue. Accordingly, researchers commonly draw great attention to this domain by discovering new venues for continuous improvements. Intuitively, KGs, as being a powerful tool for various applications, have been constructed to enrich several subdomains of Finance (TONG et al., 2016). The following discussion sheds the light on recent works pertaining to two important subdomains of Finance, namely financial investments and fraud detection.

4.5.1. Financial investments

Liu et al. (2019a) leverage a domain-specific KG to carry out stock market forecasting on the renowned companies. Their work also comprised a deep learning approach and proven effectiveness when integrated with the constructed KG on the prediction task. Another attempt was commenced by Fu et al. (2018). The authors introduced a stochastic optimization algorithm, genetic programming, and generalised crowding which are all integrated into a model for market return prediction using financial KG. Cheng et al. (2020) proposed KG-based event embedding framework that is designed for event-driven quantitative investment. In particular, the constructed KG (named FinKG) and the implemented embeddings performs learn informative representations based on both the relations of event argument and the lead-lag relations amongst the entire KG. Liu et al. (2019b) demonstrated the use of a KG embedding framework to predict stock prices using news sentiment analysis. Although the authors did not provide much discussion on the validity of the mechanism followed to construct the KG, the utility was demonstrated in the prediction task. In the same line of research, Long et al. (2020) integrated trading data, public market information and investor's records to construct a KG that is incorporated to model the market and its features. The KG was then embedded using node2vector approach and used in a deep neural network model for forecasting trends in stock prices. Authors of Ding et al. (2016) depicted the relevance of using a knowledge-empowered model on event representation and stock prediction. Zhang et al. (2018a) supported the aforementioned endeavours by proposing an approach to detect short-term stock price movement. The authors developed an enterprise KG and designed a top-up power vector model and influence propagation model. The aim was to compute the effect of a specific relationship from the relevant enterprise. The construction of the KG involved incorporating Named Entities Recognition (NER) and Neural Relation Extraction (NRE) for entities extraction and Convolutional Neural Network (CNN) for relation inference. Stock management and stock prediction tasks have been also discussed in (Wang et al., 2019c, 2020b; Deng et al., 2019b; Matsunaga et al., 2019; Zhang et al., 2018b).

4.5.2. Fraud detection

Financial fraud detection is an important area of financial risk management. KGs have been leveraged to establish approaches that can be used to stop such criminal activity. For example, Wang et al. (2020c) used a finance KG as a basis for label propagation algorithm to detect online fraud. Their model embodies a partition algorithm that is used to distinguish fraudulent groups of users. They argued that fraudulent users tend to position a close distance, whereas normal users commonly exist in isolated tense or firmly connected groups. Zhan et al. (Zhan and

---

Yin, 2018) also proposed a model that was applied in the fraud detection domain. In particular, the authors designed a call network KG that is enriched with call historical data and loan transactional data. Although the mechanism followed to construct the KG and the evaluation metrics are inadequate and inferior, the research topic is important and emphasises the significance of combining KGs with machine learning techniques to detect fraud in finance (Wang and Zhu, 2020; Kunlin, 2018).

4.6. Society and politics

Societies were initially formed men who have intuitively established political systems by making formal and informal decisions concerning the production, distribution, and the use of various resources (Sahlins, 1960). Thus, entities in terms of people, organizations, and resources have been involved to establish various forms of relations which exist among them (Schwabe et al., 2019). The formal representation of KGs provides an excellent mean to conceptualise relationships in social sciences and politics.

4.6.1. Social science

The current advances in information and communication technologies have made a qualitative leap and have created new venues where people can exchange thought, ideas and interest (Abu-Salih et al., 2018, 2020a, 2021a; Meneghello et al., 2020). This ICT revolution is embodied into various electronic means depicted by the emergence of social media. The data created by such platforms are propagating posing questions on the quality of the data being generated by these platforms (Abu-Salih et al., 2019; Wongthongtham et al., 2018; Wongthongtham and Salih, 2018; Nabipourshiri et al., 2018). Hence, there is a vital need to study these platforms and provide ground truth of trustworthy data sets (Wongthongtham and Salih, 2018; Chan et al., 2018; Abu-Salih et al., 2021b; Wongthongtham and Abu-Salih, 2015). Tchechmedjiev et al. (2019) introduced ClaimsKG, a knowledge graph of fact-checked claims originated from International Fact-Checking Network (IFCN). The purpose of ClaimsKG is to enable users searching for true facts of a certain entity. Similarly, it can be used to infer false facts of people, organizations, etc. Nguyen et al. (Nguyen and Jung, 2019) created a KG that embodied social events decomposed from social media using Independent Component Analysis (ICA) and the SocioScope Knowledge Graph (SKG) model. ICA is used to cluster social events obtained from a matrix of collected hashtags. This was followed by using the SKG model do automatically construct event-driven KGs from Twitter data. Huang et al. (2017b) reported a KG that can be used in social media to detect entity morphs (aliases that are commonly used to conceal the identity of a certain entity). The developed KG includes the real entity linked with all identified morphs mansions. A topic modelling algorithm(i.e. CorrLDA2), as well as SVM models, were used in the KG construction process. Various comparative studies were undertaken to demonstrate the effectiveness of the proposed approach.

4.6.2. Politics

Processing social data to infer domain knowledge that can be used to design a political KG was also discussed in the literature. For example, our previous work (Abu-Salih et al., 2020b) developed a credibility-based Politics and applied various KG embedding techniques to validate the KG's utility. In particular, BBC politics ontology was incorporated and extended as a backbone schema for the Politics KG. Then, various domain knowledge inference tools were used to enrich the KG with political entities captured from the collected datasets. Finally, several KG embedding models were implemented and tested over a set of link prediction, clustering, and visualisation tasks. The work proposed by Laufer and Schwabe (2017) crossed with the former model as both

efforts emphasized the significance of adding trust aspect in the process of designing and constructed a KG. In particular, the authors presented POLARE, an ontology that conceptualised the political system, and built a KG based on the provided ontology schema so as to be used for a better understanding the existing relations between agents in the political system in Brazil. Another attempt to detect and infer Political ideology was proposed by Chen et al. (2017) whereby the authors introduced an opinion-aware KG that was used for conducting political ideology forecasting. The model integrated knowledge captured from social media, DBpedia and ideological books corpus. Rudnik et al. (2019) made use of Wikidata to semantically annotating news articles. The annotated articles were then fed into a predefined event-oriented KG that was used for semantic-based search engine. News recommendation by means of a KG was also examined in Liu et al. (2019c). The authors applied a filtering method to eliminate irrelevant relations from the currently incorporated and propagated KG. Microsoft Satori KG was used as a backbone knowledge-based and enriched with entities captured from MSN news corpus. Also, the authors introduced article topic entities and the collaborative edges as two new categories of information to be embedded in the original graph. Mehdi et al. (Ali et al., 2019) designed a KG embedding approach based on socio-scholarly KG which embodies scientific artifacts on social good. The developed system incorporated various KG embedding approaches so as to retrieve, for a given entity (i.e., publication, author, domain and venue), all related and semantically-matched entities. Other studies were further established systems over KGs to benefit news and journalism domain (Berven et al., 2020; Wang et al., 2018b; Sheu and Li, 2020).

4.6.3. Culture

Modelling culture and history of societies have been also conceptualised using domain-specific KGs. For example, Liu et al. (2020) developed a KG that depicts ancient Chinese history and culture. The author constructed the KG employing Baidu Encyclopedia[18] as the knowledge source, BiLSTM-CRF for entity recognition, and DeepKE (developed by Zhejiang University) for relation inference. Constructing cultural knowledge bases that benefit from domain ontology and cultural KGs was elaborated further in Wei and Liu (2019); Marchand et al. (2020); Carriero et al. (2019).

4.7. Travel

Travel is one of the key domains which availed from the growing use of KGs. In particular, Tourism and Trafic/Transportation are amongst the subdomains that have largely employed KGs.

4.7.1. Tourism

KGs have been constructed and used in touristic QA systems (Yang et al., 2020) or in tourism recommender systems to recommend personalized attractions (Lu et al., 2016) and the best accommodation prices (Kärle et al., 2018). For example, Kärle et al. (2018) gathered data about Tirol region at Austria and constructed "Tirol Tourism" KG to benefit applications such as eCommerce. The constructed KG was fed by entities and relations extracted from Destination Management Organizations (DMOs) and Geographical Information Systems (GIS) and other tourism and accommodation-related websites. Establishing a touristic KG for China was introduced in Zhang et al. (2019b). The authors designed the domain-specific KG by using data obtained from Chinese encyclopedia KG and unstructured web pages. For entity alignment, the authors made use of Skip-Gram Model to fitch relative entities during the knowledge acquisition phase. Another attempt to construct a KG for Chinese tourism was undertaken by Yang et al. (2020). The aim was to build a QA system based on the implemented KG of tourism. The authors followed a proposed entity recognition algorithm for entity extraction

---

[18] https://baike.baidu.com/.

and utilised CNN model for relation inference. In a different context, Calleja et al. (2018) created "DBtravel" KG over the Spanish entries of Wikitravel.[19] The authors followed GATE[20] pipeline that encompasses three internal processes, namely: (1) tokenizer, (2) sentence splitter and (3) named entity recognition. Employing KGs in the tourism domain has been highly active recently and used in various applications (Feng, 2020; Liang et al., 2020; Wu et al., 2020).

4.7.2. Transportation and traffic

Applying KGs in transportation and traffic has also obtained much attention recently due to the population growth, air pollution, and other sophisticated embedded issues that require intelligent systems to resolve them efficiently (Díaz et al., 2020). For example, Zhou et al. (Zhou and Chen, 2019) introduced a model to predict urban traffic congestion. In this model urban KG was constructed from miscellanies static and dynamic raw urban data. The authors applied CNN to model spatio-temporal correlation between each indicated region. One of the well-known KGs that was applied to the US transportation system is ATMGRAPH (M. Keller, 2019) which was built at the top of the NASA ATM Ontology (ATMONTO) (Keller, 2018). M. Keller (M. Keller, 2019) in ATMGRAPH combined an array of structured aviation data obtained from the large part by US federal agencies. The developed KG conceptualises the US National Airspace System by incorporating entities describing, airspace infrastructure, flights, and flight operating conditions. Detecting traffic events by employing KG was proposed in Muppalla et al. (2017). The authors built a KG named ITSKG (Imagery-based Traffic Sensing Knowledge Graph) that was used to comprehend traffic patterns based on stationary traffic camera imagery data. Further, Wang et al. (2019d) followed a semi-automated KG construction approach based on China railway electrical accidents data. The purpose of the work is to analyse and identify the faulty equipment of railway electrical accidents in China as well to infer patterns and trends from such data. In a different context, Zhang et al. (2020c) have expanded the exertions in the maritime transportation by applying KG to build a knowledge representation of the regulations depicted by International Maritime Dangerous Goods Code (IMDG Code). They aimed to facilitate the access and retrieval of detailed guidelines embedded in IMDG. Knowledge-based models in transportation domain were also benchmarked (Chaves-Fraga et al., 2020) and others were used in travel-related intelligent systems to facilitate information sharing and interoperability (Sandberg et al., 2020; Ramnani et al., 2018).

4.8. Summary

The aim of this study was to review the latest publications concerned with KG construction approaches in seven domains of knowledge, namely: healthcare, education, ICT, science and engineering, finance, society and politics, and travel. A close examination of these interesting domains reveals important areas of research that can greatly benefit from KG technology. This study not only demonstrates the popularity of incorporating KGs to arrive at solutions for real-life problems, but also illustrates how KGs have proven to be an effective overall solution to mitigate complexity, ensure flexibility, and establish a common-ground topology whereby data can be integrated from different sources. This is clearly articulated in the selected articles in which there is a consensus that KG technology facilitates the integration of information captured from various sources. For example, we observe several efforts where KGs have been constructed using heterogenous resources to build consolidated QA and recommender systems in healthcare (Mohammadhassanzadeh et al., 2018; Huang et al., 2017a; Haussmann et al., 2019), to discover relations in biodata (Choi and Lee, 2019), to predict ecotoxicological effects in environment engineering (Myklebust et al., 2019), to

detect fraud in finance (Zhan and Yin, 2018), and to discover and visualise political relationships (Abu-Salih et al., 2020b; Chen et al., 2017). These attempts demonstrate how knowledge can be obtained from different sources, which perhaps exist in different formats, and can be then fed into one coherent schema to be formally used to conceptualise the designated domain. At the same time, we observe that several other works, particularly in the sciences and engineering domain, have not taken advantage of this important feature of KGs, but have employed limited data sources to feed their constructed KGs. Further observations and findings from this study are discussed in the next section.

5. Findings from the survey

The review of KG construction approaches which are drawn from academic works in seven domains reveals a correlated array of limitations and deficiencies related to the following summarised points:

5.1. KG data quality, privacy, and credibility

The reviewed articles have not consistently used standardized and appropriate data quality measures, particularly with the construction of large-scale KGs. Various approaches imported data collected from noisy and low-quality data sources (such as social media) with little consideration given to the credibility or privacy of generated information. This is also seen in electronic medical records from which it is difficult to collect data due to privacy and confidentiality constraints. This raises a question regarding the quality and robustness of KGs that are constructed from such data repositories. Despite some attempt to tackle data quality issues (Chen et al., 2019a; Zaveri et al., 2016), the endeavours in this direction have been inadequate. Blockchain technology presents a promising breakthrough in this regard as it provides continuous verification and advanced audit trails of all transactions. Blockchain technology does not only provide a complete mechanism to ensure data quality, integrity, and accuracy; it also has a secure and tamper-proof architecture for data storage, which would be extremely beneficial for healthcare information. Future research should be steered toward designing methodologies that establish best practices to construct *immutable* domain-specific KGs leveraging both KG and blockchain technologies.

5.2. Knowledge resources and semantic expansion

Semantic Web technologies and Linked Open Data (LOD) have opened the door wide to improving several domain applications (Hyvönen, 2012). KGs are extensions to these efforts and are commonly associated with LOD projects as they enrich the semantics of data by providing conceptual representations of concepts and entities (Ehrlinger and Wöß, 2016). Therefore, the interoperability of information is facilitated by relevant entity interlinking obtained from other KG repositories, thereby constructing multi-modal KGs. However, the approaches examined in the designated domains have demonstrated limitations in achieving interoperability of information. In particular, semantic expansion/broadening techniques were insufficiently incorporated to benefit from the openly available vocabularies and curated semantic repositories. Essentially, the underlying structure of KGs is designed to pave the way for data integration, unification, and information sharing and usability. Research in this direction should be reinforced to ensure that the essence of KGs adheres to FAIR (Findable, Accessible, Interoperable, Reusable) principles.[21]

5.3. KG construction algorithms

Deriving meaningful knowledge from a diversity of data formats is

---

[19] https://wikitravel.org/.
[20] https://gate.ac.uk/.

[21] https://www.go-fair.org/fair-principles/.

not a trivial task; it involves extracting facts about entities and their potential relationships, which requires a correlated array of various Information Extraction (IE) techniques and sophisticated Natural Language Processing (NLP) approaches. The review of the selected articles revealed only limited discussions on techniques used for entity recognition and/or relation extraction. Most of these studies either neglected to specify the algorithm(s) used in the KG construction including techniques for entity and relation extraction (e.g. (Yang et al., 2019b; Fu et al., 2018; Wang et al., 2020c)) or presented few details and poor rationale for using such techniques (e.g. (Gatta et al., 2017; Jia et al., 2018; Krinkin et al., 2020; Cheng et al., 2020)). From the analysis, we also identified an existing research gap - there were no consolidated methodologies for automating the process of constructing a domain-specific KG that facilitates the selection of suitable algorithms for the designated techniques (involving both NLP methods and data-driven approaches (Hunter, 2017)). Such methodologies would not only improve KG construction practices; they could also make the knowledge broadly available for access by both humans and machines.

5.4. Time-aware KGs

The dynamic nature of knowledge is highly correlated to contextual situations; hence, various facts that describe entities might change over time. Therefore, the temporal dimension should be integrated into KGs. Various currently propagated KGs are static but temporary, and do not consider the time factor (Jiang et al., 2016). This applies to both open-world and domain-specific KGs. Neglecting the dynamic nature of knowledge harms the quality and correctness of facts contained in the KGs and might lead to poor decision making that is based merely on such data sources. Consequently, despite some exceptions such as Wikidata and YAGO in which certain facts are already endowed with the time information, the construction of KGs should consider the validity period of facts (Leblay and Chekol, 2018).

5.5. KG evaluation

During the process of constructing KGs, incorrect facts in terms of entities and/or relations could be captured. This process is prone to errors, particularly in regard to the information derived from mix-quality data sources. As mentioned previously, it is vital to consider the credibility of the data source, as is the evaluation of the overall construction process. In fact, among the reviewed articles, the evaluation of KGs was the predominant weakness. For example, some studies carried out a superficial and subjective evaluation of the KG construction with no incorporation of concrete evaluation metrics (Fan et al., 2020b; Fu et al., 2018; Muppalla et al., 2017). Another thread of efforts attempted to involve theoretically-proven evaluation metrics to systematically measure KG completion and KG correctness approaches. The former approaches were commonly measured using recall, precision, and f-measure (Huang et al., 2019; Chi et al., 2018a; Su and Zhang, 2020; Jia et al., 2018). The latter incorporated accuracy and/or area under the ROC curve (AUC) (Paulheim, 2017; Chen et al., 2020a; Long et al., 2020). Nevertheless, our study endorses the work conducted by Paulheim (2017) in that, providing holistic techniques which simultaneously improve the quality of KGs in dissimilar domains, is still an open research problem. Further, the absence of testbeds and benchmark datasets has prevented the community from undertaking an appropriate and fair evaluation of techniques used for KG creation (Chaves-Fraga et al., 2019). Lastly, in their evaluation methodologies, researchers should consider various other data quality indicators including completeness, believability, relevance, objectivity, consistency, understandability, etc. (Wang and Strong, 1996).

5.6. Computing performance in big KGs

There is a notable consensus amongst the research community that

the conventional technologies used to process and analyse the continuous propagation large-scale datasets are no longer adequate (Marz and Warren, 2015). Machine Learning and Artificial Intelligence have become the preferred methods for the processing and analysis of Big data, by means of which the hoped-for added value is obtained. Large-scale KGs (i.e., those with trillions of triples) adhere to various Big data features. Volume is not the only feature of Big KGs; they can be described by the variety in data sources, the velocity of data generation, veracity of data quality, volatility of data currency and availability, etc. Despite some growing efforts to integrate large-scale KGs in Big Data processing (Wang and Luo, 2018), there is a need for a deeper and congruent integration of Big Data technology infrastructures and sophisticated statistical models required for reasoning over domain-specific KGs, and this remains an open research venue.

5.7. Domain-specific KG reasoning

KG reasoning aims to provide new interpretations and conclusions from constructed KGs. It is a mechanism whereby new facts can be inferred from an existing KG (Tari, 2013). KG embedding-based approaches (including tensor decomposition, distance, and semantic matching models) have gained considerable attention in the literature due to their scalability and efficiency in accommodating large-scale KGs so as to deduce a generalizable context about the KG that can be applied to infer new relations (Wang et al., 2017a; Chen et al., 2020b). This study discloses a lack of incorporating KG embedding techniques into the examined approaches. Applying KG embeddings extends these efforts and tackles the deficiencies that might occur in the KG construction process, which subsequently leads to incomplete graphs. To address this issue and to establish a technical ground for conducting relational reasoning in AI systems, KG embeddings should be integrated and implemented in domain-specific KG construction models.

5.8. Availability of domain-specific KGs

Unlike open-world and generic KGs, most of the constructed and populated domain-specific KGs are not published on the Web and thus are not accessible to other researchers. The paucity of such annotated and enriched data silos makes the process of benchmark comparison more difficult, as well as limiting reusability, interlinking and interoperability. At the same time, there is a scarcity of relevant experts who can perform data annotation. For example, in the healthcare domain, there is a lack of detailed clinical entities and relations. This is due to the shortage of both clinical repositories (Malik et al., 2020) and clinicians to conduct data annotation (Xue and Chuah, 2019). This issue has also been highlighted by researchers in Cybersecurity (Piplai et al., 2020), Finance (Bi et al., 2020), and Social Science (Feng and Chen, 2021), to name a few.

5.9. Room for further research on domain-specific KGs

This study attempts to capture a snapshot of recent advances in KGs in certain domains of knowledge. Due to limitations of space, not all KGs construction attempts in every domain and its subdomains could be considered. Hence, there are still various specific areas of the selected high-level domains that can exploit KG technology. For example, there is a need for technical solutions to respond to COVID-19 as a critical catalyser of various technological applications. KG technology can provide a shared conceptualised structure which can be used to map all dispersed facts stored in various data silos which are relevant to this pandemic. Integrating information from various available resources pertaining to COVID-19 can help to bridge the knowledge gap and provide a better understanding of this disease. COVID-19 has also given impetus to the adoption of new technologies for remote and online education. The propagation of online learning resources has established a space where domain-specific KGs can be employed to alleviate the

problem of selecting relevant resources to benefit both teachers and learners. KG technology can be applied to solve an array of problems in any domain where the data is collected from different sources. In these applications, KG technology is an excellent means of achieving interoperability, unification, interlinking, and data integration.

6. Conclusion

The widespread use and the prevalence of the Internet have led to the rapid increase in data volume which has necessitated the development of advanced data analytics that are capable of handling the propagation and the heterogeneity of such data. Knowledge graphs continue to dominate as a distinctive form of data representation and knowledge inference, and are core activity of several industrial applications. The great amount of interest in this technology is due to its underlying structure that is built on a formal conceptual representation that is depicted by a domain ontology. Therefore, domain-specific KGs have

been constructed and used to tackle several real-life problems in dissimilar domains. Yet, to date, there has been no consensual definition of a domain-specific KG. Moreover, the current methods used to construct and evaluate domain knowledge graphs are far from perfect.

This paper is the first to present an inclusive definition of domain-specific KG. Further, it provides an in-depth analysis of the current state-of-the-art knowledge graph construction research in seven domains of knowledge. The efforts undertaken in each research domain are discussed, and the shortcomings and limitations of these efforts are considered, along with future research opportunities which we hope will motivate researchers in this specific area.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Tables

Table A.1
Overview of KG approaches in Healthcare domain.

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Gatta et al. (2017)	Generic	Medical process mining	Markov Models and Careflow Mining,	Event logs	N/A	Case study	●Inadequate evaluation, ●inadequate discussion on the KG construction approach
Mohammadhassanzadeh et al. (2018)	Generic	QA	Plausible reasoning	BioASQ, DrugBank, Disease Ontology, and SemMedDB	N/A	Domain expert's verification	●Insufficient evaluation, ●evaluating the performance of query rewriting algorithm does not exist
Rotmensch et al. (2017)	Diseases	Identification of diseases and symptoms across the collected medical records	LR, NB and Bayesian network modeling	Custom, GHKG	N/A	Precision and Recall	●Inadequate to infer correct causal relations, ●concept extraction requires further elaboration
Huang et al. (2017a)	Diseases	QA - Depression disorder	NLP tools	PubMed, DrugBank, DrugBook, andUMLS	N/A	Use cases	●Lack of proper evaluation, ● insufficient use of other important medical repositories
Chen et al. (2020a)	Diseases	Benchmark comparison-Diseases and symptoms	LR, NB, and (noisy OR)	BIDMC, GHKG	N/A	AUC, F1 score, and AUPRC	● Limited data sources, ● poor causal inference methods
Ma et al. (2019)	Diseases	Episodic memory - Inductive learning	Manual integration	GDELT and ICEWS	Tucker, RESCAL, HolE, ComplEx, DistMult, ConT and Tree	MRR, and AUPRC	● Poor generalization due to the timestamps not observed in the training dataset, ● KG embeddings are prone to overfitting (too many parameters)
Huang et al. (2019)	Healthcare management	Healthy diet recommendation	CRF, SVM and DT	Healthcare websites	N/A	Precision, Recall, and F1-score	● Limited to food and dietary, ● Chinese language only, ● inadequacy to prove utility in link prediction and other knowledge discovery tasks
Haussmann et al. (2019)	Healthcare management	Food recommendation, and QA systems	Lexical similarity and string matching	DBpedia, USDA, Recipe1M and FoodOn KG	word2vec, and FastText	Case study and F1-score	● Lack of evaluation on both KG constructions and incorporated embeddings techniques, ● inadequate meaningful representations for food recommendation
Chi et al. (2018a)	Healthcare management	Food recommendation	NLP tools, CRF, SVM, NB,	China Food Composition	N/A	Precision, recall, F1-measure, and Questionnaire	● Limited data sources, ● lack of proper recommendations, *(continued on next page)*

Table A.1 (*continued*)

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
			LSTM, and KNN	and Online health websites,			• directed solely for the Chinese context

Table A.2

Overview of KG approaches in Education domain.

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Chen et al. (2018b)	Teaching and classroom resources	K-12 Education	CRF and probabilistic association rule mining	Chinese curriculum standards of mathematics	N/A	AUC	• Insufficient evaluation, • narrow scope of the proposed KG and its applications
Su and Zhang (2020)	Teaching and classroom resources	Learning assessment and recommendation	Bootstrapping construction strategy and BERT-BiLSTM-CRF	Subject teaching resources, Baidu Encyclopedia, and DBPedia	N/A	Precision, Recall and F1 measure	• Limited demonstration on the utility of the constructed KG including assessment of student learning, etc.
Zhao et al. (2019)	Teaching and classroom resources	Solving high school mathematical exercises	Complex, Triangle, Conic and Solid	Crowdsourcing and domain experts	N/A	Accuracy, Precision, Recall and F1 measure	• Limited resources used for KG construction, • limited targeted audience
Chen et al. (2018a)	Teaching and classroom resources	Learning assessment	RNN and probabilistic association rule mining	Pedagogical data and learning assessment data	N/A	AUC, Precision, Recall and F1 measure	• Lack of elucidating the effects of KG in other settings, • their schema is relatively hard to be provided for conducting benchmark comparison
Aliyu et al. (2020)	Education management	QA and course allocation scheduling	Adhoc	Structural educational information system	N/A	Case study	• Poor evaluation measures, limited KG resources, narrow scope
Wang (2020)	Education management	Internal policy control conceptualization and visualisation in higher education	Adhoc	CNKI database	pTransE	mean Silhouette	• Limited data sources, • limited application scope, poor evaluation metrics, • KG embedding was not properly demonstrated and evaluated
Yao et al. (2020)	Educational Technologies	Link Prediction	Adhoc	Knowledge Forest, Wikipedia	TransE and BERT	Mean Rank and Hits@10	• Insufficient structural and literal embedding models were used
Chi et al. (2018b)	Educational Technologies	Decision-making in academia	NLP tools, SVM, NB, and LR	Web of Science, Engineering Village, and EBSCO	N/A	F-score	• The limited scope of KG (can be expanded to include instructors and their metadata), • no domain ontology is provided as a base for the proposed KG, • limited evaluation measures

Table A.3

Overview of KG approaches in ICT domain.

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Jia et al. (2018)	Cybersecurity	Cyberattack detection	Stanford NER	Enterprise data and security websites	N/A	Precision, Recall, and F1	• Narrow KG construction approaches,

Table A.3 (*continued*)

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Deng et al. (2019a)	Cybersecurity	QA and RS for education	Adhoc and NLP tools	Wikipedia	N/A	Case study and survey	• limited evaluation with current state-of-the-art KGs in the designated domain • Limited data sources, • no benchmark comparison, • no backbone ontology schema
Kiesling et al. (2019)	Cybersecurity	Vulnerability Assessment and Intrusion Detection	Adhoc and RML Rules[1]	NVD and security online sites	N/A	Case studies	• No benchmark comparison, • inadequate rationale on the construction approach
Nayak et al. (2020)	Software development	Test cases extraction	CPT and CRF	Software documents, requirement statements and test reports	FastText algorithm	Accuracy, Precision, Recall, F1	• Transfer learning can replace the incorporated NER model, thereby using pre-trained datasets instead, • Inadequate validation to the collected requirement statements and past test reports
Schindler et al. (2020)	Software development	Investigating Software Usage in the Social Sciences	bi-LSTM and bi-LSTM-CRF	DBpedia and PLoS	N/A	Case study	• Narrow to PLoS which affected the target domain, • evaluation is inadequate as no benchmark comparison was undertaken, • the automatic linking process with DBpedia requires a further scrutiny
Fu et al. (2019)	Software development	Info. Retrieval (IT Crowdsourcing services)	Adhoc	StackOverflow, Wikipedia and crowdsourcing platforms	N/A	MRR, P@K, Recall	• Ranking modeling system can be enhanced with incorporating neural networks
Li et al. (2019)	Software development	Industrial software design and development processes	Adhoc	Generic(public databases and unstructured data sources)	N/A	Case study	• Limited domain-based data sources, • unable to provide recommendation in complex and domain-specific situations, • poor retrieval performance
Aumayr et al. (2019)	Telecom	Telecom incidents managements	NLP tools	Network incident documents	N/A	Case study on information reduction and discovery	• No discussion on the effectiveness of the incorporated algorithms/tools for KG construction, • no benchmark comparison with similar state-of-the-art
Krinkin et al. (2020)	Telcom	Television operator network monitoring	Adhoc	Data obtained by monitoring systems	N/A	Case study	• Limited discussion on KG construction and propagation, • no rationale on using the designated ontology schema, inadequate evaluation method
Xie et al. (2020)	IoT	Bridging gaps between e-devices in IoT	Adhoc	oneM2M[2]	N/A	Case study	• Inadequate discussion on concept and relation extraction approaches, • limited in data sources and application scope

[1] https://github.com/carml/carml.
[2] https://www.onem2m.org/component/rsfiles.

Table A.4
Overview of KG approaches in Science and Engineering domain.

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Farazi et al. (2020c)	Chemistry	Combustion chemistry modelling	J-Park Simulator	Linked open data	N/A	Query and simulation systems	• Limited data sources, • lack of human and machine−interaction tools
Choi and Lee (2019)	Biology	Biodata relational discovery	Manually	PubMed, CTD, BioGRID, and MalaCards	TransE, PTransE, TransR and TransH	Hits@10	• Poor KG construction approach, • unsatisfactory evaluation of the incorporated KG embedding models
Zhu et al. (2017)	Geology	Geological IR system	HanLP[1] and association rule analysis	Baike.com	N/A	Case study	• Limited data sources., undefined underlying structure, • inadequate evaluation to the utility of KG.

Table A.4 (*continued*)

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Wang et al. (2018a)	Geology	Chinese geology Knowledgebase	CRF	Geology dictionary and TCCGMR	N/A	Case study	• Limited data sources, inadequate concept and relation extraction, • no benchmark comparison with similar state-of-the-art
Yan et al. (2020)	Manufacturing engineering	Intelligent manufacturing equipment recommendation	CRF	Equipment-related data (e.g. Baidu Encyclopedia, etc.)	N/A	Case study (Information richness and effectiveness)	• Limited to level of stand-alone equipment, • scattered manufacturing data that harden data acquisition, poor accuracy
Li et al. (2020c)	Design engineering	Evolutionary Smart Product– Service System Development	Domain experts and NLP toolkits	Smart PSS prototype, user generated textual data, misc. medical websites	N/A	Showcase	• Proposed solutions for the personalized requirements are someway generic and oversimplified, • unoptimized incorporated algorithms that led to poor complexity in terms of time and space
Myklebust et al. (2019)	Environment engineering	Ecotoxicological effect prediction	LogMap	ECOTOX, NCBI and Wikidata	TransE, DistMult, and HolE	Accuracy, Precision, Recall, and F-score	• Limited data sources; can be enhanced with information about species and compounds
Fan et al. (2020b)	Electrical engineering	Improving power dispatching process.	BiLSTM-CRF	Power dispatching texts	N/A	Subjective evaluation	• Small, static and non-diversified dataset, • poor evaluation mechanism
Yang et al. (2019b)	Electrical engineering	Improving utilization of power assets information	Adhoc	PMS and ERP	N/A	Case study	• Limited evaluation measures, • limited data sources, • limited KG structure and application
Huang et al. (2020)	Oil and gas	Intelligent earch engine for oil and gas	BiLSTM-CRF	Center of Oil and Gas	N/A	Precision and Recall	• Limited data sources, • lack of optimization to the embedded search engine.

[1] https://javalibs.com/artifact/com.hankcs/hanlp.

Table A.5

Overview of KG approaches in Finance domain.

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Liu et al. (2019a)	Investment	Stock market prediction	Adhoc	Tomson Reuters, CNN	word2vec, TransE, Neural network	Accuracy and F1-score	• Limited resources led to insignificant scale of the trained dataset, • vague KG construction approach (no algorithm is indicated), • reporting the utility of the KG embedding was inadequate
Fu et al. (2018)	Investment	Market return prediction	Adhoc	Shanghai Stock Exchange and WIND Financial Terminal	N/A	Subjective evaluation based on 29 component stocks	• KG construction in terms of entities and relations was not properly addressed, • missing KG schema, • poor evaluation to the resultant KG
Cheng et al. (2020)	Investment	Event-driven quantitative investments	OpenIE v5.1	Financial news websites	N/A	Micro − F1, Weighed − F1	• Poor discussion on KG construction mechanism/ algorithm, • lack of justification on the use of customized evaluation metric
Liu et al. (2019b)	Investment	Stock price volatility prediction	Rule-based named entity recognition	Financial news websites and UQER	TransR	Accuracy (prediction task)	• Large scale KG that can affect the performance of TransR embedding model, • the KG construction validity was not scrutinized prior using it in the prediction task
Long et al. (2020)	Investment	Stock price trend prediction	Adhoc	Chinese securities companies	node2vec	AUC	• KG modelling was neither illustrated nor validated, • sentiment analysis can be integrated to obtain better performance results
Zhang et al. (2018a)	Investment	Stock price movement direction prediction	NER, NRE, and CNN	Online financial news	N/A	Acc, MCC, and FM	• Lack of temporal dimension, • limited data sources, • discussion on KG construction is inadequate
	Financial risk management	Fraud Detection	N/A	Orange Finance	N/A		• Unindicated KG construction mechanism/algorithm,

Table A.5 (*continued*)

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure (s)	Limitation(s)
Wang et al. (2020c) Zhan and Yin (2018)	Financial risk management	Fraud Detection	Adhoc	Loan transaction data and call history	word2vec	AUC, Precision, Recall and F1 measure Precision Acceleration-Recall curve	• inadequate demonstration to the KG schema • Challenging data acquisition and might lead to poor KG construction and prediction performance accordingly, • poor evaluation to the implemented KG

Table A.6

Overview of KG approaches in the Society and Politics domain.

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Tchechmedjiev et al. (2019)	Social science	QA (fact-checked claims)	TagMe tool	International Fact-Checking Network websites	N/A	Case study	• Limited data sources, • poor KG construction algorithm, limited evaluation metrics
Nguyen and Jung (2019)	Social science	Detecting and tracing social events.	SKG model	Social media (Twitter)	N/A	Precision, Recall, and F-measure	• Complexity can be improved by increasing periods and applying ICA on other cases, • limited data sources, • schema of the KG is inadequate to properly representing the relationships
Huang et al. (2017b)	Social science	Identify Entity Morphs	CorrLDA2 and SVM	DBpedia, Yago, and Freebase	N/A	Precision, recall, and F-measure	• Too many useless generate morphs can be avoided by adopting certain heuristic algorithms, • morphs can be also extended to cover not only people but events and other entities
Ali et al. (2019)	Social science	Social good recommender system	Adhoc	Academic research papers	TransE, ComplEx, TransH, TransR, TransD, DistMult and RESCAL	Mean rank and hits@k	• Limited conceptual presentation of the KG in terms of entities and relations, • results can be improved by using reinforcement learning
Abu-Salih et al. (2020b)	Politics	Link prediction, clustering, and visualisation	Adhoc using IBM Watson NLU	BBC Politics ontology, Wordnet, Google KG and light-wight ontologies.	TransE, DistMult. ComplEx, HolE, ConvE, and ConvKB,	Hits@N, MMR, Accuracy, Precision, Recall and F-score	• Limited discussion on the utility of the KG in a practical real-life example, • limited discussion on the KG construction algorithms.
Laufer and Schwabe (2017)	Politics	Trustworthy claims fact in the political system	Adhoc/ nanopublication model[1]	Se Liga na Politica (SLNP)	N/A	Case Study	• Limited direct relations between political organizations, • lack of fine-grained pat-terns in the political agent domain, • the utility of the incorporated provenance dimension was not properly validated
Chen et al. (2017)	Politics	Political ideology detection	Holistic lexicon-based approach	DBpedia, Twitter and ideological books corpus	N/A	Accuracy	• Imperfect use of evaluation strategy and metrics, • no KG Embedding was undertaken, thus, no rationale provided on the benchmark comparison, • limited data sources
Rudnik et al. (2019)	Politics	News articles retrieval	SpaCy	Wikidata	N/A	N/A	• Limited data sources,

Table A.6 (*continued*)

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Liu et al. (2019c)	Politics	Politics news recommendation	Adhoc	MSN News corpus and Microsoft Satori	TransE	AUC, NDCG @10, and F1-Score	• evaluation metrics were not provided • Sophisticated model construction that hardens the process of usability and interoperability
Liu et al. (2020)	Culture	QA and RS for Chinese ancient history and culture.	BiLSTM-CNN-CRF[2] and DeepKE	Baidu Encyclopedia	N/A	Precision, recall, and F1-score	• Inadequate use of named entity extraction techniques, • limited data sources

[1] http://nanopub.org/wordpress/.

[2] BiLSTM-CNN-CRF: Short-Term Memory-Convolutional Neural Networks-Conditions Random Field.

Table A.7
Overview of KG approaches in Travel domain

Ref.	Sub-domain	KG Usage	Construction Algorithm(s)	KG Resource(s)	Embedding Technique(s)	Evaluation Measure(s)	Limitation(s)
Kärle et al. (2018)	Tourism	Touristic IR system for Tirol at Austria	Adhoc wrapper software	DMOs, GISs, Feratel[1], Infomax[2], web maps[3], Outdooractive[4] and waldhart[5]	N/A	Case Study	• Scalability; hard to provide a wrapper for each source when it maps to Schema.org, • data collection is subject to error- prone, • poor KG construction evaluation
Zhang et al. (2019b)	Tourism	Chinese tourism-domain knowledge service	NLP -Skip-Gram Model	Sogou-T[6], Chinese Wikipedia dump[7], and Zhishi.me[8]	N/A	Accuracy	• Inadequate evaluation measures, • lack of benchmark comparisons with other related KGs
Yang et al. (2020)	Tourism	KG-QA in tourism	Adhoc entity recognition algorithm and CNN	Manual collection and NLPCC2016KBQA dataset[9]	N/A	Accuracy	• Inadequate discussion on rationale on entity and relation extraction, • limited evaluation metrics
Calleja et al. (2018)	Tourism	Spanish tourism-oriented knowledge service	Adhoc based on GATE pipeline	Wikitravel	N/A	Precision, recall and F-measure	• Construction of the KG led to import noisy data, • limited data sources.
Zhou and Chen (2019)	Transportation and traffic	Urban traffic congestion	Adhoc	Beijing traffic data and meteorological data.	N/A	Accuracy, and F1	• Limited discussion on KG construction including entity and relation extraction approaches
Muppalla et al. (2017)	Transportation and traffic	Traffic image feature extraction	Adhoc	Imagery data	N/A	Subjective evaluation	• Poor evaluation approach, • traffic cameras are prone to errors (weather, lighting, maintenance, etc.)
Wang et al. (2019d)	Transportation and traffic	Railway Electrical Accident Analysis	Adhoc and BiLSTM-CRF	China Railway electrical accidents data	N/A	Precision, Recall and F1	• Limited data sources, • inadequate discussion on utility of the constructed KG, • lack of benchmark comparison
Zhang et al. (2020c)	Transportation and traffic	Knowledge integration of maritime dangerous goods.	Adhoc	IMDG Code	N/A	Case study	• Poor discussion on entity and relation extraction algorithms, • lack of proper evaluation metrics to validate the utility of the proposed KG

[1] https://www.feratel.com/.

[2] https://www.infomax.de/.

[3] https://general-solutions.eu.

[4] https://www.outdooractive.com.

[5] https://www.waldhart.at/.

[6] https://www.sogou.com/labs/resource/t.php.

[7] https://dumps.wikimedia.org/zhwiki/.

[8] http://openkg.cn/dataset/zhishi-me-dump.

[9] https://github.com/huangxiangzhou/NLPCC2016KBQA.

References

Abu-Salih, B., Wongthongtham, P., Chan, K.Y., 2018. Twitter mining for ontology-based domain discovery incorporating machine learning. J. Knowl. Manag. 22 (5), 949–981.

Abu-Salih, B., et al., 2019. Social Credibility Incorporating Semantic Analysis and Machine Learning: A Survey of the State-Of-The-Art and Future Research Directions. Springer International Publishing, Cham.

Abu-Salih, B., et al., 2020a. Time-aware domain-based social influence prediction. J. Big Data 7 (1), 10.

Abu-Salih, B., et al., 2020b. Relational Learning Analysis of Social Politics Using Knowledge Graph Embedding arXiv preprint arXiv:2006.01626.

Abu-Salih, B., et al., 2021a. Toward a knowledge-based personalised recommender system for mobile app development. J. Univ. Comput. Sci. 27 (2), 208–229.

Abu-Salih, B., Wongthongtham, P., Zhu, D., Chan, K.Y., Rudra, A., 2021b. Social Big Data Analytics: Practices, Techniques, and Applications. Springer, Singapore.

Al-Moslmi, T., et al., 2020. Named entity extraction for knowledge graphs: a literature overview. IEEE Access 8, 32862–32881.

Ali, M., et al., 2019. Improving access to science for social good. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer.

Aliyu, I., Kana, A., Aliyu, S., 2020. Development of knowledge graph for university courses management. Int. J. Educ. Manag. Eng. 10 (2), 1.

Aumayr, E., Wang, M., Bosneag, A.-M., 2019. Probabilistic knowledge-graph based workflow recommender for network management automation. In: 2019 IEEE 20th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM). IEEE.

Bellomarini, L., Sallinger, E., Vahdati, S., 2020. Knowledge graphs: the layered perspective. In: Knowledge Graphs and Big Data Processing. Springer, Cham, pp. 20–34.

Berven, A., et al., 2020. A knowledge-graph platform for newsrooms. Comput. Ind. 123, 103321.

Bi, Y., et al., 2020. DCDIR: a deep cross-domain recommendation system for cold start users in insurance domain. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Bianchi, F., et al., 2020. Knowledge Graph Embeddings and Explainable AI arXiv preprint arXiv:2004.14843.

Bonatti, P.A., et al., 2019. Knowledge graphs: new directions for knowledge representation on the semantic web (dagstuhl seminar 18371). In: Dagstuhl Report. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Bordes, A., et al., 2013. Translating embeddings for modeling multi-relational data. In: Advances in Neural Information Processing Systems.

Bucher, B., et al., 2020. Conciliating perspectives from mapping agencies and Web of data on successful European SDIs: toward a European geographic knowledge graph. ISPRS Int. J. Geo-Inf. 9 (2), 62.

Calleja, P., et al., 2018. DBtravel: a tourism-oriented semantic graph. In: International Conference on Web Engineering. Springer.

Carlson, A., et al., 2010. In: Toward an Architecture for Never-Ending Language Learning. Aaai, Atlanta.

Carriero, V.A., et al., 2019. ArCo: the Italian cultural heritage knowledge graph. In: International Semantic Web Conference. Springer.

Catherine, R., Cohen, W., 2016. Personalized recommendations using knowledge graphs: a probabilistic logic programming approach. In: Proceedings of the 10th ACM Conference on Recommender Systems.

Chan, K.Y., et al., 2018. Affective design using machine learning: a survey and its prospect of conjoining big data. Int. J. Comput. Integrated Manuf. 33 (7), 645–669.

Chaves-Fraga, D., et al., 2019. What are the parameters that affect the construction of a knowledge graph?. In: OTM Confederated International Conferences" on the Move to Meaningful Internet Systems. Springer.

Chaves-Fraga, D., et al., 2020. GTFS-Madrid-Bench: a benchmark for virtual knowledge graph access in the transport domain. J. Web Semant. 65, 100596.

Chen, W., et al., 2017. In: Opinion-aware Knowledge Graph for Political Ideology Detection. IJCAI.

Chen, P., et al., 2018a. KnowEdu: a system to construct knowledge graph for education. Ieee Access 6, 31553–31563.

Chen, P., et al., 2018b. An automatic knowledge graph construction system for K-12 education. In: Proceedings of the Fifth Annual ACM Conference on Learning at Scale.

Chen, H., et al., 2019a. A practical framework for evaluating the quality of knowledge graph. In: China Conference on Knowledge Graph and Semantic Computing. Springer.

Chen, Y., et al., 2019b. Time-aware smart object recommendation in social Internet of things. IEEE Internet Things J. 7 (3), 2014–2027.

Chen, I.Y., et al., 2020a. Robustly extracting medical knowledge from EHRs: a case study of learning a health knowledge graph. In: Pac Symp Biocomput. World Scientific.

Chen, X., Jia, S., Xiang, Y., 2020b. A review: knowledge reasoning over knowledge graph. Expert Syst. Appl. 141, 112948.

Cheng, D., et al., 2020. Knowledge graph-based event embedding framework for financial quantitative investments. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Chi, Y., et al., 2018a. Knowledge management in healthcare sustainability: a smart healthy diet assistant in traditional Chinese medicine culture. Sustainability 10 (11), 4197.

Chi, Y., et al., 2018b. Knowledge graph in smart education: a case study of entrepreneurship scientific publication management. Sustainability 10 (4), 995.

Choi, W., Lee, H., 2019. Inference of biomedical relations among chemicals, genes, diseases, and symptoms using knowledge representation learning. IEEE Access 7, 179373–179384.

Collins, J.B., Clark, D., 2014. Towards an Ontology of Physics. NAVAL RESEARCH LAB WASHINGTON DC.

Cui, J., Yu, S., 2019. Fostering deeper learning in a flipped classroom: effects of knowledge graphs versus concept maps. Br. J. Educ. Technol. 50 (5), 2308–2328.

Cui, L., et al., 2020. DETERRENT: knowledge guided graph attention network for detecting healthcare misinformation. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Deng, Y., et al., 2019a. Knowledge graph based learning guidance for cybersecurity hands-on labs. In: Proceedings of the ACM Conference on Global Computing Education.

Deng, S., et al., 2019b. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In: Companion Proceedings of the 2019 World Wide Web Conference.

Dettmers, T., et al., 2018. Convolutional 2d knowledge graph embeddings. In: Thirty-Second AAAI Conference on Artificial Intelligence.

Díaz, G., et al., 2020. An Intelligent Transportation System to control air pollution and road traffic in cities integrating CEP and Colored Petri Nets. Neural Comput. Appl. 32 (2), 405–426.

Ding, X., et al., 2016. Knowledge-driven event embedding for stock prediction. In: Proceedings of Coling 2016, the 26th International Conference on Computational Linguistics: Technical Papers.

Dong, X., et al., 2014. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Dupre, J., 2014. A process ontology for biology. Philosopher's Mag. (67), 81–88.

Ehrlinger, L., Wöß, W., 2016. Towards a definition of knowledge graphs. SEMANTiCS (Posters, Demos, SuCCESS) 48, 1–4.

Elnagar, S., Weistroffer, H.R., 2019. Introducing knowledge graphs to decision support systems design. In: EuroSymposium on Systems Analysis and Design. Springer.

Erkimbaev, A.O., et al., 2019. Ontological concepts and taxonomies for nano world. J. Inf. Knowl. Manag. 18, 1950014, 02.

Fader, A., Soderland, S., Etzioni, O., 2011. Identifying relations for open information extraction. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.

Fan, Y., et al., 2017. Dkgbuilder: an architecture for building a domain knowledge graph from scratch. In: International Conference on Database Systems for Advanced Applications. Springer.

Fan, C., et al., 2019. Disaster City Digital Twin: a vision for integrating artificial and human intelligence for disaster management. Int. J. Inf. Manag. 102049.

Fan, R., et al., 2020a. Deep learning-based named entity recognition and knowledge graph construction for geological hazards. ISPRS Int. J. Geo-Inf. 9 (1), 15.

Fan, S., et al., 2020b. How to Construct a Power Knowledge Graph with Dispatching Data? Scientific Programming, p. 2020.

Farazi, F., et al., 2020a. Linking reaction mechanisms and quantum chemistry: an ontological approach. Comput. Chem. Eng. 106813.

Farazi, F., et al., 2020b. OntoKin: an ontology for chemical kinetic reaction mechanisms. J. Chem. Inf. Model. 60 (1), 108–120.

Farazi, F., et al., 2020c. Knowledge graph approach to combustion chemistry and interoperability. ACS Omega 5 (29), 18342–18348.

Feng, L., 2020. Design of tourism intelligent recommendation model of mount tai scenic area based on knowledge graph. In: 2020 International Conference on E-Commerce and Internet Technology (ECIT). IEEE.

Feng, D., Chen, H., 2021. A small samples training framework for deep Learning-based automatic information extraction: case study of construction accident news reports analysis. Adv. Eng. Inf. 47, 101256.

Fu, X., et al., 2018. Stochastic optimization for market return prediction using financial knowledge graph. In: 2018 IEEE International Conference on Big Knowledge (ICBK). IEEE.

Fu, D., et al., 2019. In: Enhancing Semantic Search of Crowdsourcing IT Services Using Knowledge Graph. SEKE.

Fujun, Z., Quanhui, Y., Luyun, Y., 2017. Analysis of technical opportunity in marine science field based on knowledge graph. Sci. Technol. Manag. Res. 2017 (24), 25.

Gao, J., et al., reportEfficient Knowledge Graph Accuracy Evaluation (Technical Report Version).

Gao, X., et al., 2019. Construction of knowledge map of continuous fiber reinforced ceramic matrix composites. In: Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition.

Gatta, R., et al., 2017. Generating and comparing knowledge graphs of medical processes using pMineR. In: Proceedings of the Knowledge Capture Conference.

Goodwin, T.R., Harabagiu, S.M., 2017. Knowledge representations and inference techniques for medical question answering. ACM Trans. Intell. Syst. Technol. (TIST) 9 (2), 1–26.

Groth, P., Gibson, A., Velterop, J., 2010. The anatomy of a nanopublication. Inf. Serv. Use 30 (1–2), 51–56.

Hastings, J., et al., 2012. Structure-based classification and ontology in chemistry. J. Cheminf. 4 (1), 8.

Haussmann, S., et al., 2019. FoodKG: a semantics-driven knowledge graph for food recommendation. In: International Semantic Web Conference. Springer.

Heist, N., 2018. Towards knowledge graph construction from entity Co-occurrence. EKAW (Doctoral Consortium).

Heist, N., et al., 2020. Knowledge Graphs on the Web-An Overview.

Hooi, E.K.J., et al., 2019. TAGraph: knowledge graph of threat actor. In: 2019 International Conference on Cybersecurity (ICoCSec). IEEE.

Huang, Z., et al., 2017a. Constructing knowledge graphs of depression. In: International Conference on Health Information Science. Springer.

Huang, L., et al., 2017b. KIEM: a knowledge graph based method to identify entity morphs. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.

Huang, L., et al., 2019. Towards smart healthcare management based on knowledge graph technology. In: Proceedings of the 2019 8th International Conference on Software and Computer Applications.

Huang, S., Wang, Y., Yu, X., 2020. Design and implementation of oil and gas information on intelligent search engine based on knowledge graph. In: Journal of Physics: Conference Series. IOP Publishing.

Hubauer, T., et al., 2018. Use cases of the industrial knowledge graph at siemens. In: International Semantic Web Conference (P&D/Industry/BlueSky).

Humayun, F., et al., 2020. A computational approach for mapping heme biology in the context of hemolytic disorders. Front Bioeng. Biotechnol. 8, 74.

Hunter, L.E., 2017. Knowledge-based biomedical data science. EPJ Data Sci 1 (1–2), 19–25.

Hyvönen, E., 2012. Publishing and using cultural heritage linked data on the semantic web. Synth. Lect. Semant. Web: Theor. Technol. 2 (1), 1–159.

Ji, S., et al., 2020. A Survey on Knowledge Graphs: Representation, Acquisition and Applications arXiv preprint arXiv:2002.00388.

Jia, Y., et al., 2018. A practical approach to constructing a knowledge graph for cybersecurity. Engineering 4 (1), 53–60.

Jiang, T., et al., 2016. Towards time-aware knowledge graph completion. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers.

Jiang, B., et al., 2019. Intelligent interaction with virtual geographical environments based on geographic knowledge graph. ISPRS Int. J. Geo-Inf. 8 (10), 428.

Jiménez-Ruiz, E., Grau, B.C., 2011. Logmap: logic-based and scalable ontology matching. In: International Semantic Web Conference. Springer.

Kärle, E., et al., 2018. Building an ecosystem for the tyrolean tourism knowledge graph. In: International Conference on Web Engineering. Springer.

Kejriwal, M., 2019. Domain-Specific Knowledge Graph Construction. Springer.

Kejriwal, M., Shao, R., Szekely, P., 2019. Expert-guided entity extraction using expressive rules. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Keller, R., 2018. *The NASA Air Traffic Management Ontology (Atmonto)–release Dated March 2018*. 2019, Technical Report. URL …. National Aeronautics and Space Administration.

Kiesling, E., et al., 2019. The SEPSES knowledge graph: an integrated resource for cybersecurity. In: International Semantic Web Conference. Springer.

Kim, K., et al., 2020. GREG: a global level relation extraction with knowledge graph embedding. Appl. Sci. 10 (3), 1181.

Kipf, T.N., Welling, M., 2016. Semi-supervised Classification with Graph Convolutional Networks arXiv preprint arXiv:1609.02907.

Krdzavac, N., et al., 2019. An ontology and semantic web service for quantum chemistry calculations. J. Chem. Inf. Model. 59 (7), 3154–3165.

Krinkin, K., et al., 2020. Models of telecommunications network monitoring based on knowledge graphs. In: 2020 9th Mediterranean Conference on Embedded Computing (MECO). IEEE.

Kuhn, P., et al., 2016. Type inference on Wikipedia list pages. Informatik 2016.

Kunlin, Y., 2018. A memory-enhanced framework for financial fraud detection. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE.

Laufer, C., Schwabe, D., 2017. On modeling political systems to support the trust process. In: PrivOn@ ISWC..

Le-Phuoc, D., et al., 2016. The graph of things: a step towards the live knowledge graph of connected things. J. Web Semant. 37, 25–35.

Leblay, J., Chekol, M.W., 2018. Deriving validity time in knowledge graph. In: Companion Proceedings of the the Web Conference 2018.

Li, R., et al., 2019. A knowledge graph framework for software-defined industrial cyber-physical systems. In: IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society. IEEE.

Li, Y., et al., 2020a. Domain specific knowledge graphs as a service to the public: powering social-impact funding in the US. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Li, L., et al., 2020b. Real-world data medical knowledge graph: construction and applications. Artif. Intell. Med. 103, 101817.

Li, X., et al., 2020c. A knowledge graph-aided concept–knowledge approach for evolutionary smart product–service system development. J. Mech. Des. 142 (10).

Li, Z., et al., 2020d. Demonstration of fault localization in optical networks based on knowledge graph and graph neural network. In: Optical Fiber Communication Conference. Optical Society of America.

Liang, C., et al., 2018. In: Investigating Active Learning for Concept Prerequisite Learning. AAAI.

Liang, X., Cao, H., Zhang, W., 2020. Knowledge extraction experiment based on tourism knowledge graph Q & A data set. In: 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS). IEEE.

Lin, Z., Yang, D., Yin, X., 2020. Patient Similarity via Joint Embeddings of Medical Knowledge Graph and Medical Entity Descriptions. IEEE Access.

Liu, H., et al., 2016. Learning concept graphs from online educational data. J. Artif. Intell. Res. 55, 1059–1090.

Liu, Y., et al., 2019a. Anticipating stock market of the renowned companies: a knowledge graph approach. Complexity, 2019.

Liu, J., Lu, Z., Du, W., 2019b. Combining enterprise knowledge graph and news sentiment analysis for stock price prediction. In: Proceedings of the 52nd Hawaii International Conference on System Sciences.

Liu, D., et al., 2019c. News graph: an enhanced knowledge graph for news recommendation. In: KaRS@ CIKM.

Liu, S., et al., 2020. Preliminary study on the knowledge graph construction of Chinese ancient history and culture. Information 11 (4), 186.

Long, J., et al., 2020. An integrated framework of deep learning and knowledge graph for prediction of stock trend: an application in Chinese stock exchange market. Appl. Soft Comput. 106205.

Lu, C., Laublet, P., Stankovic, M., 2016. Travel attractions recommendation with knowledge graphs. In: European Knowledge Acquisition Workshop. Springer.

M Keller, R., 2019. Building a knowledge graph for the air traffic management community. In: Companion Proceedings of the 2019 World Wide Web Conference.

Ma, Y., Tresp, V., Daxberger, E.A., 2019. Embedding models for episodic knowledge graphs. J. Web Semant. 59, 100490.

Malik, K.M., et al., 2020. Automated domain-specific healthcare knowledge graph curation framework: subarachnoid hemorrhage as phenotype. Expert Syst. Appl. 145, 113120.

Marchand, E., Gagnon, M., Zouaq, A., 2020. Extraction of a knowledge graph from French cultural heritage documents. In: ADBIS, TPDL and EDA 2020 Common Workshops and Doctoral Consortium. Springer.

Marz, N., Warren, J., 2015. Big Data: Principles and Best Practices of Scalable Realtime Data Systems. Manning Publications Co.

Matsunaga, D., Suzumura, T., Takahashi, T., 2019. Exploring Graph Neural Networks for Stock Market Predictions with Rolling Window Analysis arXiv preprint arXiv: 1909.10660.

Meissner, R., Köbis, L., 2020. Annotated knowledge graphs for teaching in higher education. In: International Conference on Web Engineering. Springer.

Meneghello, J., et al., 2020. Unlocking social media and user generated content as a data source for knowledge management. Int. J. Knowl. Manag. 16 (1), 101–122.

Menon, A., Krdzavac, N.B., Kraft, M., 2019. From database to knowledge graph—using data in chemistry. Curr. Opin. Chem. Eng. 26, 33–37.

Mohammadhassanzadeh, H., et al., 2018. Investigating plausible reasoning over knowledge graphs for semantics-based health data analytics. In: 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). IEEE.

Muppalla, R., et al., 2017. A knowledge graph framework for detecting traffic events using stationary cameras. In: Proceedings of the 2017 ACM on Web Science Conference.

Myklebust, E.B., et al., 2019. Knowledge graph embedding for ecotoxicological effect prediction. In: International Semantic Web Conference. Springer.

Nabipourshiri, R., Abu-Salih, B., Wongthongtham, P., 2018. Tree-based classification to users' trustworthiness in OSNs. In: Proceedings of the 2018 10th International Conference on Computer and Automation Engineering. ACM, Brisbane, Australia, pp. 190–194.

Nayak, A., Kesri, V., Dubey, R.K., 2020. Knowledge graph based automated generation of test cases in software engineering. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, pp. 289–295.

Nguyen, H.L., Jung, J.J., 2019. Social event decomposition for constructing knowledge graph. Future Generat. Comput. Syst. 100, 10–18.

Nguyen, D.Q., et al., 2017. A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network arXiv preprint arXiv:1712.02121.

Nickel, M., Tresp, V., Kriegel, H.-P., 2011. In: A Three-Way Model for Collective Learning on Multi-Relational Data. Icml.

Nickel, M., et al., 2015. A review of relational machine learning for knowledge graphs. Proc. IEEE 104 (1), 11–33.

Nickel, M., Rosasco, L., Poggio, T., 2016. Holographic embeddings of knowledge graphs. In: Thirtieth Aaai Conference on Artificial Intelligence.

Paulheim, H., 2017. Knowledge graph refinement: a survey of approaches and evaluation methods. Semantic Web 8 (3), 489–508.

Pezeshkpour, P., Tian, Y., Singh, S., 2020. Revisiting Evaluation of Knowledge Base Completion Models. Automated Knowledge Base Construction.

Pingle, A., et al., 2019. Relext: relation extraction using deep learning approaches for cybersecurity knowledge graph improvement. In: Proceedings of the 2019 IEEE/ ACM International Conference on Advances in Social Networks Analysis and Mining.

Piplai, A., et al., 2020. Creating cybersecurity knowledge graphs from malware after action reports. IEEE Access 8, 211691–211703.

Ramnani, R.R., Sengupta, S., Debnath, P., 2018. Intelligent travel advisor: a goal oriented virtual agent with task modeling, planning and user personalization. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents.

Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Comput. Geosci. 31 (9), 1119–1125.

Rastogi, N., Zaki, M.J., 2020. Personal Health Knowledge Graphs for Patients arXiv preprint arXiv:2004.00071.

Rotmensch, M., et al., 2017. Learning a health knowledge graph from electronic medical records. Sci. Rep. 7 (1), 5994.

Rudnik, C., et al., 2019. Searching news articles using an event knowledge graph leveraged by wikidata. In: Companion Proceedings of the 2019 World Wide Web Conference.

Sadeghi, A., Lehmann, J., 2019. Linking physicians to medical research results via knowledge graph embeddings and Twitter. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer.

Sahlins, M.D., 1960. The origin of society. Sci. Am. 203 (3), 76–88.

Sahu, A., et al., 2018. Method and System for Content Processing to Query Multiple Healthcare-Related Knowledge Graphs. Google Patents.

Salehian, H., et al., 2019. Food Knowledge Graph for a Health Tracking System. Google Patents.

Sandberg, M.K., et al., 2020. Explaining traffic situations–architecture of a virtual driving instructor. In: International Conference on Intelligent Tutoring Systems. Springer.

Schindler, D., Zapilko, B., Krüger, F., 2020. Investigating software usage in the social sciences: a knowledge graph approach. In: European Semantic Web Conference. Springer.

Schwabe, D., Laufer, C., Busson, A., 2019. Building Knowledge Graphs about Political Agents in the Age of Misinformation arXiv preprint arXiv:1901.11408.

Shen, Y., et al., 2019. KGDDS: a system for drug-drug similarity measure in therapeutic substitution based on knowledge graph curation. J. Med. Syst. 43 (4), 92.

Sheng, M., et al., 2020. DSQA: a domain specific QA system for smart health based on knowledge graph. In: International Conference on Web Information Systems and Applications. Springer.

Sheu, H.-S., Li, S., 2020. Context-aware graph embedding for session-based news recommendation. In: Fourteenth ACM Conference on Recommender Systems.

Shi, D., et al., 2020. A Learning Path Recommendation Model Based on a Multidimensional Knowledge Graph Framework for E-Learning. Knowledge-Based Systems, p. 105618.

Singhal, A., 2012. Introducing the Knowledge Graph: Things, Not Strings. Google.

Smirnova, A., Cudré-Mauroux, P., 2018. Relation extraction using distant supervision: a survey. ACM Comput. Surv. 51 (5), 1–35.

Su, Y., Zhang, Y., 2020. Automatic construction of subject knowledge graph based on educational big data. In: Proceedings of the 2020 the 3rd International Conference on Big Data and Education.

Suchanek, F.M., Kasneci, G., Weikum, G., 2007. Yago: a core of semantic knowledge. In: Proceedings of the 16th International Conference on World Wide Web.

Sun, K., et al., 2016. Visualization for knowledge graph based on education data. Int. J. Softw. Inf. 10 (3).

Sun, H., et al., 2020. Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: model development and performance evaluation. JMIR Med Inform 8 (7), e17653.

Tao, X., et al., 2020a. Mining Health Knowledge Graph for Health Risk Prediction. World Wide Web, pp. 1–22.

Tao, Y., Li, M., Hu, W., 2020b. Research on knowledge graph model for cybersecurity logs based on ontology and classified protection. In: Journal of Physics: Conference Series. IOP Publishing.

Tari, L., 2013. Knowledge inference. Encycl. Syst. Biol. 1074–1078.

Taylor, K.R., et al., 2006. Bringing chemical data onto the semantic web. J. Chem. Inf. Model. 46 (3), 939–952.

Tchechmedjiev, A., et al., 2019. ClaimsKG: a knowledge graph of fact-checked claims. In: International Semantic Web Conference. Springer.

Tong, R., Xue, L., Wang, H., 2016. Building and Exploring an Enterprise Knowledge Graph for Investment Analysis. GROTH P, SIMPERL E, GRAY A, et al. The Semantic Web-ISWC.

Trouillon, T., et al., 2016. Complex embeddings for simple link prediction. In: International Conference on Machine Learning (ICML).

Wang, J., 2020. Knowledge graph analysis of internal control field in colleges. Teh. Vjesn. 27 (1), 67–72.

Wang, Y., Luo, J., 2018. An incremental reasoning algorithm for large scale knowledge graph. In: International Conference on Knowledge Science, Engineering and Management. Springer.

Wang, R.Y., Strong, D.M., 1996. Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. 12 (4), 5–33.

Wang, C., Zhu, H., 2020. Representing fine-grained Co-occurrences for behavior-based fraud detection in online payment services. IEEE Trans. Dependable Secure Comput.

Wang, Q., et al., 2017a. Knowledge graph embedding: a survey of approaches and applications. IEEE Trans. Knowl. Data Eng. 29 (12), 2724–2743.

Wang, X., et al., 2017b. Community preserving network embedding. In: Thirty-first AAAI Conference on Artificial Intelligence.

Wang, C., et al., 2018a. Information extraction and knowledge graph construction from geoscience literature. Comput. Geosci. 112, 112–120.

Wang, H., et al., 2018b. DKN: deep knowledge-aware network for news recommendation. In: Proceedings of the 2018 World Wide Web Conference.

Wang, R., et al., 2019a. Knowledge graph embedding via graph attenuated attention networks. IEEE Access 8, 5212–5224.

Wang, H., et al., 2019b. Service application knowledge graph and dependency system. In: 2019 34th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW). IEEE.

Wang, Z., Zhang, X., Hu, Y., 2019c. A semantic path based approach to match subgraphs from large financial knowledge graph. In: 2019 International Conference on Mathematics, Big Data Analysis and Simulation and Modelling (MBDASM 2019). Atlantis Press.

Wang, X., Wang, J., Han, J., 2019d. Knowledge Graph Construction for Railway Electrical Accident Analysis. In: 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). IEEE.

Wang, F., et al., 2020a. Recent advances on graph analytics and its applications in healthcare. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Wang, H., Wang, T., Li, Y., 2020b. Incorporating expert-based investment opinion signals in stock prediction: a deep learning framework. In: Proceedings of the AAAI Conference on Artificial Intelligence.

Wang, J., et al., 2020c. Improving Graph-Based Label Propagation Algorithm with Group Partition for Fraud Detection. Applied Intelligence, pp. 1–10.

Wei, J., Liu, R., 2019. A versatile approach for constructing a domain knowledge graph for culture. Proc. Assoc. Inf. Sci. Technol. 56 (1), 808–809.

Wongthongtham, P., Abu-Salih, B., 2015. Ontology and trust based data warehouse in new generation of business intelligence: state-of-the-art, challenges, and

opportunities. In: Industrial Informatics (INDIN), 2015 IEEE 13th International Conference on. IEEE.

Wongthongtham, P., Salih, B.A., 2018. Ontology-based approach for identifying the credibility domain in social Big Data. J. Organ. Comput. Electron. Commer. 28 (4), 354–377.

Wongthongtham, P., et al., 2018. State-of-the-Art ontology annotation for personalised teaching and learning and prospects for smart learning recommender based on multiple intelligence and fuzzy ontology. Int. J. Fuzzy Syst. 20 (4), 1357–1372.

Wu, W., et al., 2012. Probase: a probabilistic taxonomy for text understanding. In: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data.

Wu, J., et al., 2020. Event-centric tourism knowledge graph—a case study of hainan. In: International Conference on Knowledge Science, Engineering and Management. Springer.

Xiao, H., et al., 2019. Embedding and predicting software security entity relationships: a knowledge graph based approach. In: International Conference on Neural Information Processing. Springer.

Xie, R., et al., 2019. DeepLink: a code knowledge graph based deep learning approach for issue-commit link recovery. In: 2019 IEEE 26th International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE.

Xie, C., et al., 2020. Multi-layer Internet of things middleware based on knowledge graph. IEEE Internet Things J.

Xinqing, W., et al., 1999. APPLICATION OF" THE TERMINOLOGY CLASSIFICATION CODES OF GEOLOGY AND MINERAL RESOURCES, vol. 5. ON GEOLOGICAL AND MINERAL RESOURCES POINT-SOURCE INFORMATION SYSTEM.

Xue, Q., Chuah, M.C., 2019. Explainable deep learning based medical diagnostic system. Smart Health 13, 100068.

Yan, H., Yang, J., Wan, J., 2020. KnowIME: a system to construct a knowledge graph for intelligent manufacturing equipment. IEEE Access 8, 41805–41813.

Yang, B., et al., 2014. Embedding Entities and Relations for Learning and Inference in Knowledge Bases arXiv preprint arXiv:1412.6575.

Yang, A., et al., 2019a. Enhancing pre-trained language representations with rich knowledge for machine reading comprehension. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.

Yang, Y., et al., 2019b. Multi-source heterogeneous information fusion of power assets based on knowledge graph. In: 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI). IEEE.

Yang, L., et al., 2020. Research on tourism question answering system based on Xi'an tourism knowledge graph. In: Journal of Physics: Conference Series. IOP Publishing.

Yao, S., et al., 2020. Joint embedding learning of educational knowledge graphs. In: Artificial Intelligence Supported Educational Technologies. Springer, pp. 209–224.

Yu, H., et al., 2020. A domain knowledge graph construction method based on Wikipedia. J. Inf. Sci., 0165551520932510

Yuan, J., et al., 2020. Constructing biomedical domain-specific knowledge graph with minimum supervision. Knowl. Inf. Syst. 62 (1), 317–336.

Zaveri, A., et al., 2016. Quality assessment for linked data: a survey. Semantic Web 7 (1), 63–93.

Zeng, J., Nakano, Y.I., 2020. Exploiting a large-scale knowledge graph for question generation in food preference interview systems. In: Proceedings of the 25th International Conference on Intelligent User Interfaces Companion.

Zhan, Q., Yin, H., 2018. A loan application fraud detection method based on knowledge graph and neural network. In: Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence.

Zhang, N., Zhang, L., 2020. A Knowledge Graph Analysis on PE Teachers' Research Hotspots. Design Engineering, pp. 352–366.

Zhang, N., et al., 2016. Semantic framework of Internet of things for smart cities: case studies. Sensors 16 (9), 1501.

Zhang, Y., et al., 2018a. In: Predicting Stock Price Movement Direction with Enterprise Knowledge Graph. PACIS.

Zhang, X., et al., 2018b. Improving stock market prediction via heterogeneous information fusion. Knowl. Base Syst. 143, 236–247.

Zhang, X., et al., 2019a. Research and analysis on the field of food additive by knowledge graph construction. In: Journal of Physics: Conference Series. IOP Publishing.

Zhang, W., et al., 2019b. The Chinese knowledge graph on domain-tourism. In: Advanced Multimedia and Ubiquitous Engineering. Springer, pp. 20–27.

Zhang, X., et al., 2020. Spatiotemporal features based geographical knowledge graph construction. SCIENTIA SINICA Informationis 50 (7), 1019–1032.

Zhang, Y., et al., 2020a. HKGB: an Inclusive, Extensible, Intelligent, Semi-auto-constructed Knowledge Graph Framework for Healthcare with Clinicians' Expertise Incorporated. Information Processing & Management, p. 102324.

Zhang, L., et al., 2020b. Construction and applications of embedded aerospace software defect knowledge graph. In: Signal and Information Processing, Networking and Computers. Springer, pp. 470–477.

Zhang, Q., et al., 2020c. Construction of knowledge graph of maritime dangerous goods based on IMDG code. J. Eng. 2020 (13), 361–365.

Zhao, Y., Smidts, C., 2019. A method for systematically developing the knowledge base of reactor operators in nuclear power plants to support cognitive modeling of operator performance. Reliab. Eng. Syst. Saf. 186, 64–77.

Zhao, T., et al., 2019. Towards automatic mathematical exercise solving. Data Sci. Eng. 4 (3), 179–192.

Zheng, Y., Liu, R., Hou, J., 2017. The construction of high educational knowledge graph based on MOOC. In: 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE.

Zhou, G., Chen, F., 2019. Urban congestion areas prediction by combining knowledge graph and deep spatio-temporal convolutional neural network. In: 2019 4th

International Conference on Electromechanical Control Technology and Transportation (ICECTT). IEEE.

Zhu, Y., et al., 2017. Intelligent Learning for Knowledge Graph towards Geological Data. Scientific Programming, 2017.

Zhu, Y., et al., 2020. Knowledge-driven drug repurposing using a comprehensive drug knowledge graph. Health Inf. J. 1460458220937101.

Zou, X., 2020. A survey on application of knowledge graph. JPhCS 1487 (1), 012016.

Bilal Abu-Salih: is an Assistant Professor at the University of Jordan and an Adjunct at Curtin University. He holds a Ph.D. in Information Systems (with a focus on Social Big Data Analytics) from Curtin University since 2018. He worked with various cross-disciplinary funded research projects which are related to data analytics, machine learning, data mining of social media, big data analysis etc. Those projects are involved with academic research and also involved software development and industrial implementation. Bilal's research interests include; Knowledge Graphs, Social Big Data, Social Trust, Machine Learning/Data Mining, Semantic Analytics, NLP, Information Retrieval, and the like.