

Received January 19, 2021, accepted March 16, 2021, date of publication April 1, 2021, date of current version April 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070395

Knowledge Graphs in Manufacturing and Production: A Systematic Literature Review

GEORG BUCHGEHER¹, DAVID GABAUER¹, JORGE MARTINEZ-GIL¹,
AND LISA EHRLINGER^{1,2}

¹Software Competence Center Hagenberg GmbH, 4232 Hagenberg, Austria

²Institute for Application-Oriented Knowledge Processing (FAW), Johannes Kepler University Linz, 4040 Linz, Austria

Corresponding author: Georg Buchgeher (georg.buchgeher@scch.at)

This work was supported in part by the Interreg Österreich-Bayern 2014-2020 Programme funded under Grant (AB292), in part by the Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK), in part by the Federal Ministry for Digital and Economic Affairs (BMDW), and in part by the Province of Upper Austria in the frame of the COMET – Competence Centers for Excellent Technologies Programme managed by Austrian Research Promotion Agency FFG.

ABSTRACT Knowledge graphs in manufacturing and production aim to make production lines more efficient and flexible with higher quality output. This makes knowledge graphs attractive for companies to reach Industry 4.0 goals. However, existing research in the field is quite preliminary, and more research effort on analyzing how knowledge graphs can be applied in the field of manufacturing and production is needed. Therefore, we have conducted a systematic literature review as an attempt to characterize the state-of-the-art in this field, i.e., by identifying existing research and by identifying gaps and opportunities for further research. We have focused on finding the primary studies in the existing literature, which were classified and analyzed according to four criteria: bibliometric key facts, research type facets, knowledge graph characteristics, and application scenarios. Besides, an evaluation of the primary studies has also been carried out to gain deeper insights in terms of methodology, empirical evidence, and relevance. As a result, we can offer a complete picture of the domain, which includes such interesting aspects as the fact that knowledge fusion is currently the main use case for knowledge graphs, that empirical research and industrial application are still missing to a large extent, that graph embeddings are not fully exploited, and that technical literature is fast-growing but still seems to be far from its peak.

INDEX TERMS Knowledge graphs, manufacturing, production, systematic literature review.

I. INTRODUCTION

The twenty-first century has been clearly marked by its rapid growth in artificial intelligence (AI) applications. Thus, companies are required to undergo an inherent transformation to leverage AI for reaching Industry 4.0 standards and for gaining a competitive advantage in the international market. While AI technologies such as neural networks, natural language processing, chatbots, autonomous driving vehicles, and digital twins received increasing attention in the field of manufacturing and production, little light is shed on the applications of knowledge graphs (KGs) in this domain.

In recent years, a large number of open (public) as well as closed (enterprise) KGs have been developed. While open KGs, which are often academic and open-source projects, provide access to anyone on the web, enterprise KGs are

closed applications within companies that are only accessible to approved users [1]. Examples for public KG projects are DBpedia [2], Freebase [3], KBpedia [4], NELL [5], PROSPERA [6], Wikidata [7], and YAGO [8]. The most popular commercial and closed KGs are Cyc [9], Google Knowledge Graph [10], [11], Google Knowledge Vault [12], and Microsoft Satori. The blog entry by Google [10]¹ is frequently quoted as seminal work of KG research since it sparked the discussion in this field in 2012. Considering early KG research from the 1980s, Google has rather revived KG technology than invented it. The foundation of KGs has been laid out by Sowa [13], who provided conceptual graph theory as an early stage contribution for knowledge representation in semantic networks [14]. Further seminal work on KGs has been conducted by Stokman and co-authors, who aimed at building a KG to represent medical and sociological

The associate editor coordinating the review of this manuscript and approving it for publication was Shunfeng Cheng.

¹Note that Google's blog entry has been recently updated in [11].

literature [15]–[17].² Even though knowledge graphs are nowadays frequently applied in different domains, there is still no formal definition, which is accepted in the entire community. In 2016, Ehrlinger and Wöß [18] proposed the following widely-acknowledged definition:

Definition Knowledge Graph: A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge.

Ehrlinger and Wöß [18] further discuss existing alternative definitions and their implications and limitations. To be in-line with the state-of-the-art, our paper covers all studies that claim to employ a knowledge graph if (1) they have been accepted by a scientific peer-reviewed journal or conference, or if (2) they conform to at least one definition reviewed by Ehrlinger and Wöß. In summary, the common denominator of a KG is its structure in terms of nodes (entities) and edges (relationships). For storing graphs, the two most popular data models are *RDF triples* and *property graphs*.³ The majority of public KGs is stored in form of Resource Description Framework (RDF) triples (subject-predicate-object), proposed by the World Wide Web Consortium (W3C).⁴ In RDF, subjects and objects are nodes and predicates the edges between the nodes. Property graphs store nodes and edges natively, whereas the nodes can have properties in form of key-value pairs [19].

KGs are primarily used to semantically model a specific and often complex domain [20]. This explicitly modeled domain knowledge is used to support and enhance the accuracy of downstream tasks like question answering [21], [22], information extraction [23], [24], named entity disambiguation [25], [26], semantic parsing [27], [28], and recommender systems [29], [30]. Also, the analysis of KGs with machine learning methods, e.g., to predict missing edges or to classify nodes, has gained increasing attention [31]. Since most machine learning models require a set of feature vectors as input, much research has been done to generate “embeddings” from KGs. A KG embedding transforms the nodes and (depending on the approach) also the edges to a numeric feature vector [32], which serves as direct input to a machine learning model. Considering the plethora of application scenarios mentioned above, several domains have already perceived the substantial benefits KG technology brings with it. Example domains, which already rely on the use of KGs, are science [33], healthcare [34]–[37], cybersecurity [38]–[41], data defects [42], education and training [43]–[45], and tourism [46].

To lay out the foundation for novel research in using KGs in the field of manufacturing and production, we conducted a systematic literature review. The aim of this review is to

²The increasing popularity of knowledge graphs led to the funding of several projects related to knowledge graphs in manufacturing and other domains. See <https://www.nsf.gov/od/oia/convergence-accelerator/Award%20Listings/track-a.jsp>.

³We refer to [19] for a discussion on less common graph data models, e.g., hypergraphs, which are not further considered in the frame of this paper.

⁴<https://www.w3.org/RDF>

systematically capture the current state-of-the-art of KGs by revealing their utilization in manufacturing and production. Already available standards and consensus with respect to the structure and construction of KGs in this domain should be discovered and discussed. By having an overview of the current state-of-the-art, existing solutions can be refined and extended, as well as novel approaches can be developed that have not yet been the subject of research. According to the OECD,⁵ more than 70 % of the G7’s world trade is based upon goods. Even though less than 25 % of all jobs are provided by industries,⁶ a significant amount of jobs in other sectors depend on the jobs in the production sector. Despite the substantial size and importance of this sector, KGs have been neglected as one of the key AI technologies so far. Thus, this research contributes to the dissemination and use of KGs in industry applications by highlighting their benefits and how companies can leverage them. We aim at answering the research question: “Which role play knowledge graphs in manufacturing and production?” The question is answered by an investigation of (1) the bibliometric key facts, (2) research type facets, (3) KG characteristics, and (4) KG application scenarios.

The remainder of this paper is structured as follows: Section II presents related work and Section III describes the planning and realization of the systematic literature review. The results, which were obtained from analyzing the primary studies and answers to the research questions are provided in Section IV. In Section V, we further discuss the research questions along with open research challenges and the threats to validity. Section VI concludes our study.

II. RELATED WORK

To the best of our knowledge, there is no systematic literature review or systematic mapping study dedicated to knowledge graphs in manufacturing and production. Yet, there are still surveys, reviews, and books that aim to provide an overview on the state-of-the-art of KG technologies.

Chronologically, we start with Nickel *et al.* [47], who provide the first survey on KGs with a special focus on the usage of latent and graph feature models for retrieving knowledge to predict new facts/edges in the graph. The foundation, architecture, construction, and applications of enterprise knowledge graphs is outlined in detail by Pan *et al.* in [48]. Paulheim [49] describes how to refine a knowledge graph based upon its A-box via completion, error detection, types of refinement, internal and external methods, and puts forward various evaluation standards that can be employed. The study of Wang *et al.* [50] is similar to [47] with a comprehensive summary of translational distance and semantic matching models in the field of KG embeddings and a comment on the usefulness of KGs with respect to recommender systems and question answering applications. In [51]

⁵See, <https://data.oecd.org/trade/trade-in-goods-and-services.htm> OECD. G7 countries trade 2.323,096 million USD in services and 5.844,610 million USD in goods.

⁶See, WorldBank.

Lin *et al.* employ a subset of the KG embeddings presented in [47], [50] and address complex relation modeling, relational path modeling, and multi-source information learning. Contrary to all previously mentioned articles, Yan *et al.* [52] do not focus on a specific KG topic, but gives a general overview on how KGs are constructed.⁷ The work of Gesese *et al.* [53] is the first survey on KG embeddings that make use of literals. Furthermore, Kazemi *et al.* [54] outlined how representation learning approaches are expedient for dynamic graphs. The book of Kejriwal [55] provides a very general summary of domain-specific KG construction. One survey about fault domain knowledge graphs has been written by Wang and Yang [56]. Another important scrutiny has been conducted by Al-Moslmi *et al.* [57] outlining pre-processing tools based on natural language processing, such as, named entity recognition, named entity disambiguation, and named entity linking, to enable the construction of a KG. Fensel *et al.* [46] provide a recent introduction into knowledge graphs with a lot of well-relatable real-life examples. Heist *et al.* [58] give an overview of cross-domain KGs that are publicly available on the Web. Ji *et al.* [59] extend the study of [50] by explaining how different kinds of neural networks can be used to generate KG embeddings. Finally, a recent study by Hogan *et al.* [60] comprises all of the aforementioned studies' topics and provides a profound and comprehensive foundation into the field of knowledge graphs starting from scratch covering both, deductive and inductive knowledge representation techniques.

III. RESEARCH METHOD

According to Brereton *et al.* [61], systematic literature reviews (SLRs) are “*a means of evaluating and interpreting all available research relevant to a particular research question or topic area or phenomenon of interest*”. SLRs are secondary empirical studies used to provide a structured overview of a research field [62]. An SLR follows a well-defined methodology, which makes it less likely that the results of the literature are biased, although it does not protect against publication bias in the primary studies [62].

In this SLR, we followed the steps outlined in [62] – a systematic literature review consists of three main phases, i.e., the planning of the SLR, conducting the SLR, and reporting the SLR. This section presents the planning of the study, i.e., the research questions, the data sources, and search strategy, along with the classification and evaluation criteria.

A. RESEARCH QUESTIONS

The aim of this SLR is to analyze the current status of knowledge graphs in the field of manufacturing and production. Thus, existing research is investigated to identify potential

⁷Keep in mind that this paper was submitted already in June 2015 and got accepted for publication in January 2016 which in turn means it has been available prior [47] which could explain the more general view on the topic matters.

TABLE 1. Research questions.

Nr.	Research questions
RQ1	What are the bibliometric key facts of KG publications?
RQ2	Which research type facets do the identified publications address?
RQ3	What are the specific knowledge graph characteristics?
RQ4	What are application scenarios of knowledge graphs?

gaps and opportunities for future work. The main research question guiding this study is:

Which role play knowledge graphs in manufacturing and production?

The research question we established for this study attempts to provide specific insights into the relevant aspects of how KGs are used in production and manufacturing. This includes questions about the articles' bibliometric key facts, research type facets, specific KG characteristics, and their application scenarios. We also want to examine the type of research carried out up to that time (theoretical, proposal, empirical), together with the type of research forums in which these works have been published and presented. The exact research questions this SLR answers are reported in Table 1.

RQ1 provides an overview of bibliometrics of published studies, concerning knowledge graph applications in production and manufacturing to exhibit the importance and timeliness of this topic. In more detail, we analyze the publication trend, publication venues, and origin countries of research institutes that have published studies in this field. RQ2 investigates the maturity of knowledge graph applications by analyzing which research methods have been used for the validation of research. The specific construction techniques of knowledge graphs are addressed in RQ3. This is of major importance for consultants and practitioners as it reveals the structure of knowledge graphs employed in a production and manufacturing setting. Finally, RQ4 examines the application scenarios in which knowledge graphs have been used in the context of production and manufacturing, i.e., in which particular manufacturing domains knowledge graphs have been used, for which concrete use cases knowledge graphs are used, and which kinds of systems are developed based on knowledge graphs.

B. DATA SOURCES AND SEARCH STRATEGY

To build an adequate search string we have selected two major search terms: ‘*Method*’ and ‘*Field*’. The first major search term represents the employed methodology, namely, ‘knowledge graph’ whereas the second major search term illustrates the field in which the knowledge graph should have been used. This term includes all sorts of technologies and synonyms of manufacturing and production in which the knowledge graph application should take place. Terms like ‘enterprise’, ‘industry’, ‘company’, ‘corporate’, ‘manufacturer’, ‘manufacturing’, ‘organization’, and ‘production’ should cover all synonyms for production and manufacturing, whereby, we included the German word ‘industrie’,

TABLE 2. Search string.

	Search terms
Method	("knowledge graph") AND
Field	((("enterprise")) OR ("industry")) OR (("industrie")) OR ("physical system") OR ("internet of things") OR ("company") OR ("corporate") OR ("organization") OR ("product") OR ("production") OR ("management") OR ("manufacturer") OR ("manufacturing"))

as well, since knowledge graphs could also be applied in terms of *Industrie 4.0* which is also commonly used in the international academic literature. Furthermore, 'internet of things' should supply us with references concerning 'internet of things' and 'industrial internet of things' whereas 'physical system' is related to literature with a focus on the combination of knowledge graphs and cyber-physical systems. In addition, 'enterprise' and 'management' retrieve references with respect to 'enterprise knowledge graphs', and finally, we specifically outlined 'product' for 'product knowledge graphs' which is a rather new but interesting field of knowledge graph applications.

The final search string that has been used in the presented study is shown in Table 2. The search terms were constructed using steps described in [61], in which the Boolean *OR* is used to incorporate alternative spellings, synonyms or related terms, and the Boolean *AND* is combining the link to major terms.

The proposed search strategy is set out in Table 3. The scope of the search considers publications and contributions presented in both academic and professional forums and publications. That is, we have considered academic publications (such as those published in journals or presented in academic conferences or peer-reviewed books) in addition to publications and contributions presented in industry or professional forums, such as conferences, workshops, and online publications. For academic publications, the sources of choice are: *ACM Digital Library*, *IEEEExplore*, *ISI Web of Science*, *ScienceDirect* and *Springer*. It has been a need to use a general search engine which in our case is *Google Scholar* to include non-academic contributions and publications. Certain criteria on the data sources has been invoked to overcome particular challenges to avoid assessing hundreds of thousands of articles. To keep the search within reasonable bounds, we restricted the number of results retrieved from Google Scholar to 300.⁸ What is more, this data source was applied only to search for non-academic primary studies: those papers

⁸Note, that this number was sufficiently high, since a significant part of the last results returned by the engine did not include any primary studies.

TABLE 3. Summary of the search strategy.

Search strategy	
Academic databases	ACM Digital Library IEEEExplore ISI Web of Science ScienceDirect Springer
Other data sources	Google Scholar
Target items	Books Conference papers Workshop papers Journal papers Industry/professional workshop contribution Industry/professional conference contribution Non-academic online publications
Search applied to	Title Abstract Keywords Full-text (Google Scholar)

or articles published in industry/professional conferences, workshops, online journals/magazines or corporate blogs. The strategic search has been conducted recursively, that is, relevant studies referenced in the primary studies will also be considered. Personal blogs or web pages have been excluded from the search.

The inclusion and exclusion criteria whether a paper is taken into account for the systematic literature review is shown in Table 4. Every study needs to include at least one of both major search terms. Additionally, it has to be published in an academic or professional forum. English has to be the language of the full-text and the publication date is not allowed to exceed the 26th of February, 2020. In case, the inclusion criteria have been fulfilled and none of the exclusion criteria has been triggered as well, the study will be considered as a primary study in the SLR.

In the first round, the title and abstract of each study are analyzed whether the paper is an eligible fit for the SLR. Although corporate blog posts are considered, personal blogs or web pages are strictly excluded. In case, the paper is only available in the form of a *PowerPoint* presentation or the emphasis of the article is not on knowledge graphs in a production or manufacturing setting it is excluded as well.

In the second round, we are left with all papers that have been affirmed to be relevant in the first round. In this round, also the full texts of the papers are considered. If an article only has an abstract but no full-text, or represents a summary of a workshop it is excluded. Non-academic or non-professional papers are eliminated as well. Further, we have dropped papers discussing knowledge graphs but only refer to manufacturing and production as potential usage domain without describing a concrete application scenario of knowledge graphs in this domain. For deciding if a paper describes a use case from the manufacturing domain we have used the North American Industry Classification Sys-

TABLE 4. Summary of the selection strategy.

Inclusion and exclusion criteria	
Inclusion criteria	IC-1: Terms fulfill the search string IC-2: Academic journal, conference and workshop papers IC-3: Contribution to industry/professional conferences, workshops, and online publications IC-4: Papers written in English IC-5: Publication date: until 26th of February, 2020
Exclusion criteria for titles and abstract	EC-1.1: Personal blogs or web pages EC-1.2: Papers available only in the form of PowerPoint presentations EC-1.3: Papers which do not focus on knowledge graphs in manufacturing or production
for full text	EC-2.1: Papers available only in the form of abstracts EC-2.2: Papers presenting a summary of a workshop EC-2.3: Non-academic/non-professional online publications EC-2.4: Papers which are using manufacturing or production applications just as examples EC-2.5: Papers which do not focus on knowledge graphs in manufacturing according to the NAICS

TABLE 5. Data items extracted from each study.

Nr.	Item name	Description	Relevant RQ
D1	Publication year	In which year was the article published?	RQ1
D2	Venue	What is the name of the journal, conference, or workshop?	RQ1
D3	Country	Where are the research institutes located that have published the study?	RQ1
D4	Research field	What are the Scimago classifications of the outlet?	RQ2
D5	Evidence level	What is the evidence level of the evaluation of the proposed approach?	RQ2
D6	KG approach	What kind of knowledge graph and which technique has been used?	RQ3
D7	Domain	In which manufacturing domain has the knowledge graph been constructed?	RQ4
D8	Use Case	Which use case is supported by the knowledge graph?	RQ4
D9	System Kind	Which kind of system has been developed based on a knowledge graph?	RQ4

tem (NAICS)⁹ as a reference. The NAICS is a classification system for business establishments providing a systematic overview of the manufacturing domain. All other articles that are in-line with the inclusion and exclusion criteria are considered as primary studies.

C. DATA EXTRACTION AND SYNTHESIS

To answer the RQs defined in Table 1, we extract specific data from the selected primary studies. Table 5 highlights the data items (D1 to D9) extracted for the analysis in this review. D1, D2, and D3 provide clues concerning the distribution of knowledge graph studies in manufacturing and production over years, venues, and countries of publication, and thus answers RQ1. D4 and D5 directly contribute to the answers of RQ2. D6 can be used to answer RQ3. D7, D8, and D9 contribute to the answer of RQ4 and further discussion of knowledge graph approaches in manufacturing and production. To ensure that the data extraction results are unbiased, two authors performed the data extraction for each primary study independently, and then one checked the data extraction

⁹According to NAICS, the manufacturing sectors are, food manufacturing, beverage and tobacco product manufacturing, textile mills, textile product mills, apparel manufacturing, leather and allied product manufacturing, wood product manufacturing, paper manufacturing, printing and related support activities, petroleum and coal products manufacturing, chemical manufacturing, plastics and rubber products manufacturing, nonmetallic mineral product manufacturing, primary metal manufacturing, fabricated metal product manufacturing, machinery manufacturing, computer and electronic product manufacturing, electrical equipment, appliance, and component manufacturing, transportation equipment manufacturing, furniture and related product manufacturing, and miscellaneous manufacturing.

results of the other, and finally, they discussed and reached a consensus on the data extraction results.

D. EVALUATION

A six-point Likert-scale was designed to provide a quality assessment of the selected primary studies. We categorize studies to five different research type facets that are weighted according to their quality of evidence (as proposed by [63]), namely from *Opinion Papers '1'* to *Evaluation Research '5'*. The final numerical value which generates the evaluation of each paper assumes a value between 0 and 1. The evaluation provides insights into the degree to which different aspects of knowledge graphs are considered in existing research in the field. It was decided that the results of this assessment would help to identify the quality of research carried out.

The questions composing the quality assessment are shown in Table 6 and follow the six-level classification of evidence evaluation suggested by [64]. The purpose of these evaluation questions was to assess the primary studies based upon the employed methodology, as well as, how the proposal has been integrated.

IV. EMPIRICAL RESULTS

In this section, we offer the detailed results of our literature analysis. Thus, this section is structured around the four research questions we have answered.

A. RESULTS OF THE SEARCH

The initial search resulted in 833 articles. A detailed break down by databases is shown in Table 7. We have obtained

TABLE 6. Evaluation level of research type facet (see [63]).

Research type facet	Evidence level	Description
Evaluation Research	1.0	Techniques are implemented in practice and an evaluation of the technique is conducted. That means, it is shown how the technique is implemented in practice (solution implementation) and what are the consequences of the implementation in terms of benefits and drawbacks (implementation evaluation). This also includes identifying problems in the industry.
Validation Research	0.8	Techniques investigated are novel and have not yet been implemented in practice. Techniques used are for example experiments.
Solution Proposal	0.6	A solution for a problem is proposed, the solution can be either novel or a significant extension of an existing technique. The potential benefits and the applicability of the solution is shown by a small example or a good line of argumentation.
Philosophical Papers	0.4	These papers sketch a new way of looking at existing things by structuring the field in form of a taxonomy or conceptual framework.
Experience Papers	0.2	Experience papers explain what and how something has been done in practice. It has to be the personal experience of the author.
Opinion Papers	0.0	These papers express the personal opinion of somebody whether a certain technique is good or bad, or how things should be done. They do not rely on related work and research methodologies.

TABLE 7. Electronic databases included in this SLR.

Nr.	Database	Number of studies
DB1	ACM Digital Library	227
DB2	IEEEExplore	172
DB3	ISI Web of Science	84
DB4	ScienceDirect	23
DB5	Springer	21
DB6	Google Scholar	356

227, 172, 84, 23, 21, and 356 studies from *ACM Digital Library*, *IEEEExplore*, *ISI Web of Science*, *ScienceDirect*, *Springer*, and *Google Scholar*,¹⁰ respectively.

Removing the duplicates of all 833 found publications leaves us left with 745 studies. Out of those 745 studies, analyzing the abstract has revealed that 682 are not relevant for the SLR. 5 out of the remaining 63 articles have not been written in English. After reading the full-text of all 58 articles, 24 have been identified to be relevant as primary studies for this systematic literature review. In each step, at least two authors needed to have the same opinion on whether a paper is included or excluded. In case, two authors had a different opinion on one study, a third author decided whether to keep the primary paper in the selection process. 2 articles have been added as a result of snowballing the references of all papers. The detailed breakdown of the search process is given in Figure 1. As the findings indicate, the number of primary studies obtained may appear to be quite small – there are just 24, however, as will be shown in greater detail in this section, all these papers were published between the years 2016 and 2020. The full list of primary studies gathered is listed in the appendix.

B. RESEARCH QUESTIONS

In this section, we answer each research question outlined in Table 1 by discussing the analysis of the primary studies.

¹⁰In the case of Google Scholar, two searches have been conducted, whereas the first search targeted all studies that have the search string included in the title of the article (56 articles), and the second search extracted the first 300 most relevant articles that include the search string in the manuscript.

1) RQ1: WHAT ARE THE BIBLIOMETRIC KEY FACTS OF KG FOR PUBLICATIONS?

Figure 2 shows the distribution of the studies according to the year they have been published. The first primary studies with a focus on KGs in a manufacturing or production environment date back to 2016. This highlights the fact that this field of research is still in its early stages of development. The number of studies published in 2017 is six times that of 2016. Moreover, the number of publications in 2017 is larger than in 2018. However, the number of primary studies published in 2019 is more than twice as much as in 2018. As we have only covered the first two months in 2020 the number of published studies can be disregarded, however, deducting from previous years we expect that KG in production and manufacturing studies will exceed 2019 as a result of its increasing popularity.

Next, we have analyzed the origin countries of the primary studies. Therefore we have analyzed the affiliations of the authors, i.e., the countries of their research institutes. For each study, we have taken the origin countries of all authors into consideration. Figure 3 depicts the results of the analysis and shows that most primary studies originate from research institutes located in Germany (8 studies), and China (7 studies). Four studies were published by research institutes from Italy, followed by three studies from the United Kingdom and the United States each. Both India and Singapore are the origin of two studies, and one study originates from Finland and Russia each.

After the analysis of the primary studies by country, we continue by focusing on the type of publication forum. The ordered distribution of the type of forums is shown in Figure 4. A closer look at Figure 4 points out that 12 primary studies – which are exactly half of all primary studies – have been published in conference proceedings. 8 primary studies – representing one-third of all studies – have found their outlet in journals. In addition, only 2 articles have been printed in workshops whereas 1 primary study has been available as a book, and 1 primary study is still a pre-print.

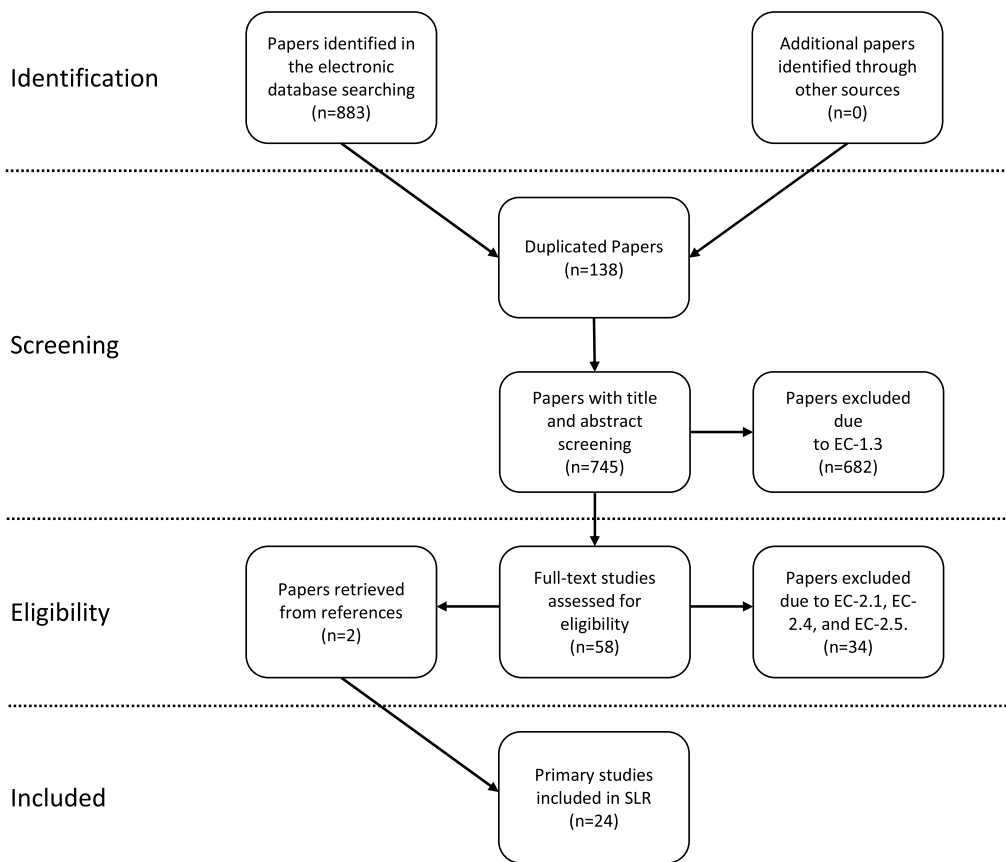


FIGURE 1. Primary study selection process.

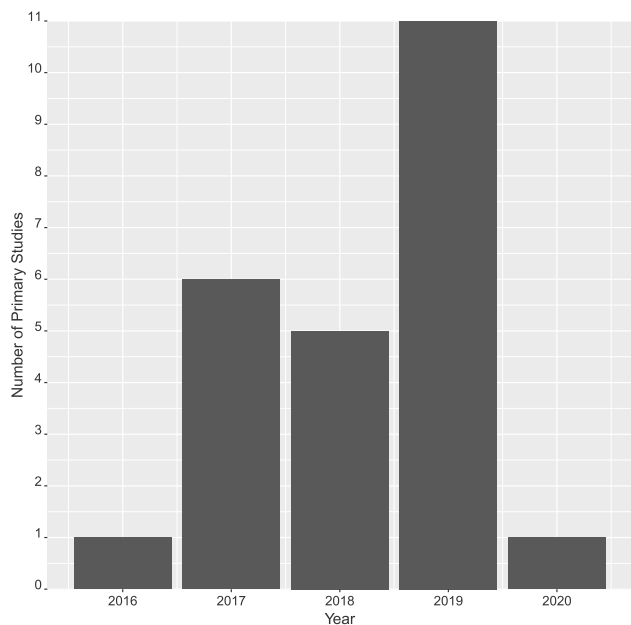


FIGURE 2. Distribution of primary studies by year.

We continue by classifying the primary studies according to their Scimago¹¹ research field. The classification by

¹¹<https://www.scimagojr.com>

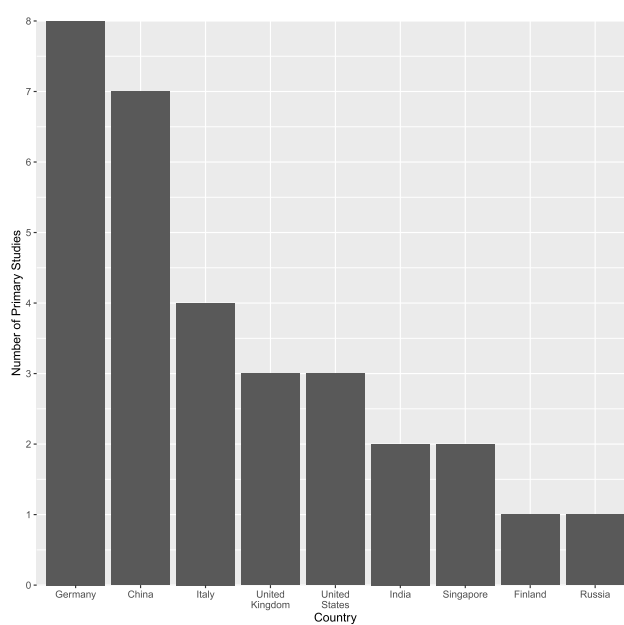


FIGURE 3. Distribution of primary studies by country.

Scimago is not unique and therefore, a single study can be assigned to multiple research fields. Figure 5 reveals that most studies have been published in *Computer Science*. In more

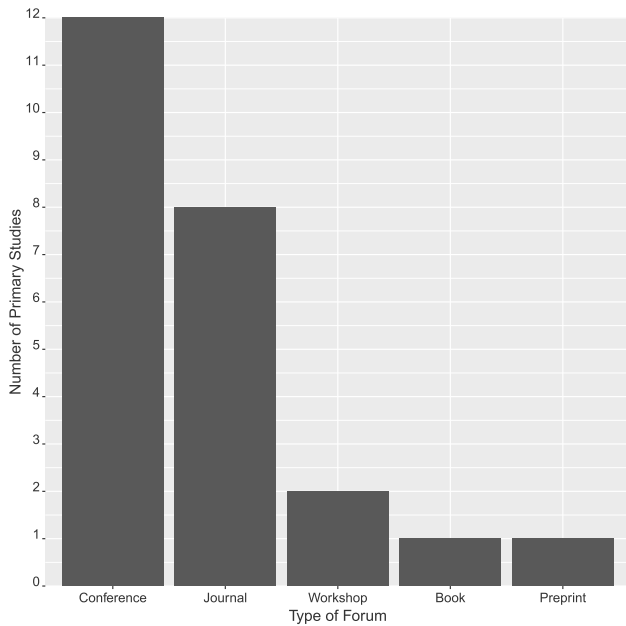


FIGURE 4. Distribution of primary studies by forum.

detail, 19 articles and hence almost 80 % of the articles belong to this category. 9 primary studies are considered in the field of *Engineering* which represent 37.5 % of all selected papers. Additionally, 3 studies have been published in outlets that are considered in the field of *Business, Management and Accounting*. Finally, all remaining fields – *Chemical Engineering, Materials Science, Mathematics, Physics, and Astronomy, and Social Science* – inherent 2 primary studies. This demonstrates that KGs mainly belong to the discipline of computer science and engineering.

2) RQ2: WHICH RESEARCH TYPE FACETS DO THE IDENTIFIED PUBLICATIONS ADDRESS?

The question regarding the research type facet is of fundamental importance as it demonstrates the applicability of the approach and whether it has already been implemented in a real-world scenario. This information is not only essential for researchers, but it is also of major concern for practitioners and consultants.

Figure 6 explicitly displays the clear picture of the primary studies by research facet type. In this regard, the majority of primary studies – 11 in total – can be considered as solution proposals since although they have paved the ground for novel KG applications, neither evaluation nor validation took place. 9 selected articles fall into the category of validation research. Hereby, the knowledge graph methodology has been employed in a real-world setting. More information about the specific application scenario is given in Section IV-B4. Less attention has been brought to evaluation research since only 2 papers have been conducted so far. In both remaining categories – experience papers and opinion papers – we have found a single study. This analysis shows that even though multiple solutions for industry-specific problems have been proposed, there is not a wealth of

TABLE 8. Classification of primary studies by research type facet.

Research type facet	Studies
Evaluation research	[S1] [S2]
Validation research	[S3] [S4] [S5] [S6] [S7] [S8] [S9] [S10] [S11]
Solution proposal	[S12] [S13] [S14] [S15] [S16] [S17] [S18] [S19] [S20] [S21] [S22]
Experience paper	[S23]
Opinion research	[S24]

literature yet that has incorporated those suggestions into real-world applications. This step is clearly missing at the current state of the literature. The details are illustrated in Table 8.

3) RQ3: WHAT ARE THE SPECIFIC KNOWLEDGE GRAPH CHARACTERISTICS?

In this section, we examine the type and construction of the employed KGs. According to literature, a KG can either be constructed by a bottom-up or a top-down approach [65], [66]. In the top-down approach, a domain expert conceptually models the schema of the KG, often in form of an ontology, which is then populated with data to complete the knowledge graph. In the bottom-up approach, the structure of the knowledge graph is (typically automatically) induced from the data [65], [66]. In terms of data model, a KG can either be an RDF graph or a property graph [19], [67]–[70].

The decision for a specific construction approach and data model clearly influences a successful application of the KG. The top-down approach, for instance, is very restrictive when adding new triples due to carefully modeled schema constraints. The bottom-up approach, on the other hand, allows rapid growth of the KG with little human effort, which could however lead to KGs with lower quality (e.g., incorrect relationships). Interestingly, we did not find many papers, which used machine learning or deep learning approaches to construct a KG.

Figure 7 shows that 69.57 % of the KGs have been constructed with a top-down approach and only 30.43 % with a bottom-up approach. This is an interesting finding since a growing amount of research focuses on the automated generation of KGs with bottom-up approaches. For example, Wang et al. [50] use neural networks to induce the structure of a KG.

With respect to the data model, Figure 8 discloses that the majority of KGs (87.5 %) are modeled as RDF graphs and only 12.5 % are modeled as property graphs. Even though 12.5 % is not much, it is still more than expected since there is a strand of literature that claims that knowledge graphs are RDF graphs [18].

4) RQ4: WHAT ARE THE APPLICATION SCENARIOS OF KNOWLEDGE GRAPHS?

We have analyzed the primary studies with regard to the application domain, use case, and system kind. Table 9 provides a summary of each primary study.

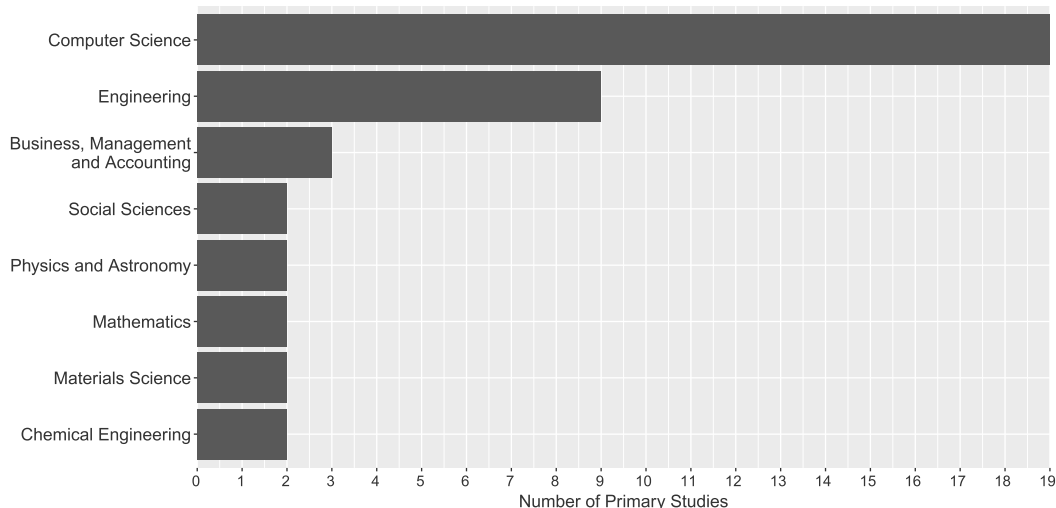


FIGURE 5. Distribution of primary studies by Scimago research field.

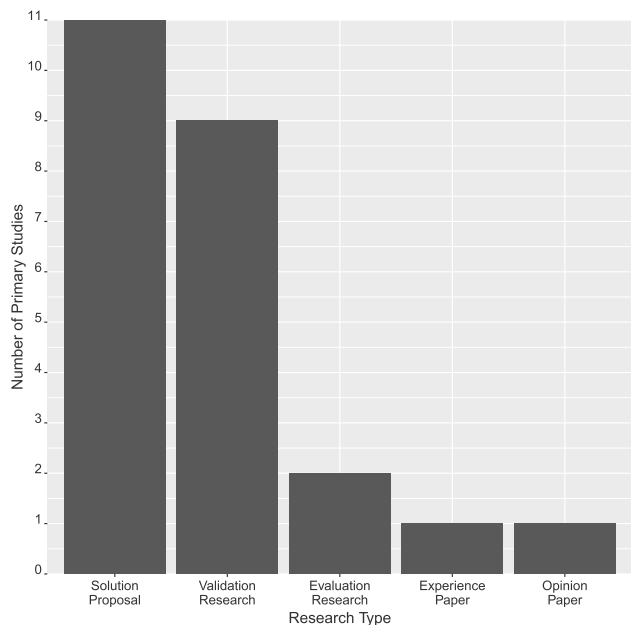


FIGURE 6. Distribution of primary studies by research type facet.

The application domain describes in which field of manufacturing and production KGs have been used. Figure 9 depicts the results of the classification with respect to the NAICS schema. As shown in the figure, in half of the analyzed primary studies - 12 primary studies - no concrete application domain was provided. 4 primary studies describe the use of KGs in the field of *machinery manufacturing*, 3 primary studies come from the field of *chemical manufacturing*, 3 primary studies come from the field of *transportation equipment manufacturing*, 1 primary study has been published in the field of *fabricated metal product manufacturing*, and 1 primary study comes from the context of *textile product*

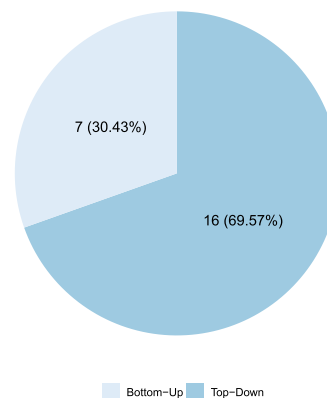


FIGURE 7. Distribution of kind of knowledge graph.

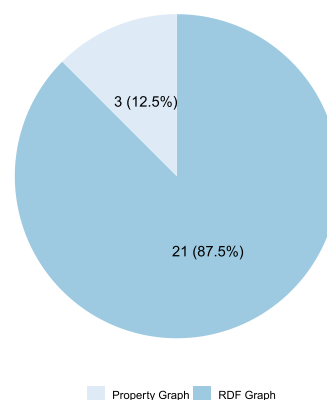


FIGURE 8. Distribution of knowledge graph construction.

mills. The assignment of each primary study to an application domain is shown in Table 10.

We have analyzed each primary study for the motivating use case, i.e., why a knowledge graph has been constructed. Figure 10 provides an overview of which use cases have been described in the selected primary studies. 50 % of

TABLE 9. Summary of primary studies.

Primary Study	Summary
[S1]	Development of a KG based on data of a fashion manufacturer. The KG integrates different knowledge sources, e.g., data of different products and sales data. On top of the KG, a data visualization tool is developed supporting dynamic data analysis and sales forecasting.
[S2]	Construction of a KG based on big data in the field of additive manufacturing as the foundation for a data analysis platform.
[S3]	Construction of a KG representing a digital twin of an automated production line. The KG integrates contextual data (e.g., equipment configuration) with operational data (process data and events, e.g., alarms). A machine learning component adapts the KG in response to changes of the manufacturing environment.
[S4]	Generation of a KG based on sensor data of a production line and inference of additional relationships based on an ontology.
[S5]	A framework for the construction of a KG based on descriptions of cyber-physical systems described using smart manufacturing standards, e.g., AutomationML.
[S6]	Construction of a KG representing a cyber-physical system based on multiple descriptions of different perspectives of the CPS, which are semantically integrated with each other.
[S7]	Definition of a KG storing manufacturing knowledge and production problems. On top of the knowledge graph, a knowledge platform is constructed. The platform provides a knowledge matching component that links production problems to related manufacturing knowledge.
[S8]	In this work, a KG is used for the generation of action plans for autonomous robots. The KG is comprised of environment data, robot capabilities, task templates, and goal descriptions.
[S9]	A general-purpose KG platform based on open standards. The platform has been evaluated in an industrial setting for integrating static and runtime data at a company from the manufacturing domain.
[S10]	Use of the J-Park Simulator for the construction of a distributed KG for the simulation of an air pollution scenario from the process industry.
[S11]	Creation of a KG from natural language requirement documents with the goal to support the automated generation of test cases from these documents. Application of the approach in the automotive domain.
[S12]	Construction of a KG representing a digital twin of an automated production line. The KG integrates contextual data (e.g., equipment configuration) with operational data (process data and events, e.g., alarms). A machine learning component adapts the KG in response to changes of the manufacturing environment.
[S13]	Construction of an KG for metallic material. The KG is constructed based on DBpedia data, which is then enriched with data from Wikipedia. Data of the KG can be displayed via a dedicated prototype.
[S14]	Construction of an open industrial KG by integrating multiple knowledge source (i.e., technical documents, data models). The KG supports the intelligent search for manufacturing services, by searching and evaluating manufacturers of machinery parts.
[S15]	Construction of an industrial KG as the foundation for a digital twin based on multiple data sources, e.g., machine data, process data, material data. A machine learning component analyzes the KG and event data (log files) to complete the KG with additional links.
[S16]	Development of a KG in the field of electrical discharge machining by integrating different data sources in a knowledge graph. Automated matching of identical concepts by applying intelligent matching algorithms that are used for the automated construction of an ontology.
[S17]	Development of a KG from CAM Models. The KG integrates the CAD models with NC process data of the CAM models and supports the interchange of different CAM Models via ontology matching.
[S18]	Use of the J-Park Simulator for the construction of a KG as foundation for the development of a smart contract-based market place. The system supports the automated search for services including service rating, service contracting, and service payment.
[S19]	Use of a KG as the foundation of single shared information space for describing objects of a manufacturing enterprise.
[S20]	Construction of a KG from unstructured Chinese text from multiple heterogeneous data sources for the automotive industry. The graph is built by applying natural language processing and deep learning for the identification of entities and relationships
[S21]	Use of a KG for automated code generation in the field of automated guided vehicles.
[S22]	Construction of a KG from isolated data silos to support consistent data provision and to improve factory planning processes.
[S23]	In this work a KG is used for the generation of action plans for autonomous robots. The KG is comprised of environment data, robot capabilities, task templates, and goal descriptions.
[S24]	Use of KG as the foundation for machine learning based on graph embeddings.

TABLE 10. Manufacturing domains of primary studies.

Domain	Studies
NA / General	[S3] [S4] [S5] [S6] [S7] [S8] [S9] [S14] [S15] [S19] [S21] [S24]
Machinery Manufacturing	[S2] [S16] [S17] [S23]
Chemical Manufacturing	[S10] [S12] [S18]
Transportation Equipment Manufacturing	[S11] [S20] [S22]
Fabricated Metal Product Manufacturing	[S13]
Textile Product Mills	[S1]

all primary studies, i.e., 12 studies, describe a knowledge fusion use case where KGs are used for integrating data from different information sources with each other. 3 primary studies are using KGs for automatically integrating separated

manufacturing processes with each other, 3 primary studies deal with the creation of digital twins based on data stored in a knowledge graph, 3 primary studies use KG data for the automated generation of source code. 1 primary study deals with the development of a general KG management platform, 1 primary study describes a predictive analytics use case. For one of the primary studies, no concrete use case could be identified. Table 11 links the assignment of each primary study to its motivating use case.

Finally, we have investigated which kinds of systems have been developed on top of a knowledge graph. Figure 11 depicts the identified kinds of systems. In 8 cases a search-based application, i.e, an application providing a search engine based on semantic technologies, has been developed. In 3 cases, a code generation system has been proposed. 2 primary studies demonstrated the development

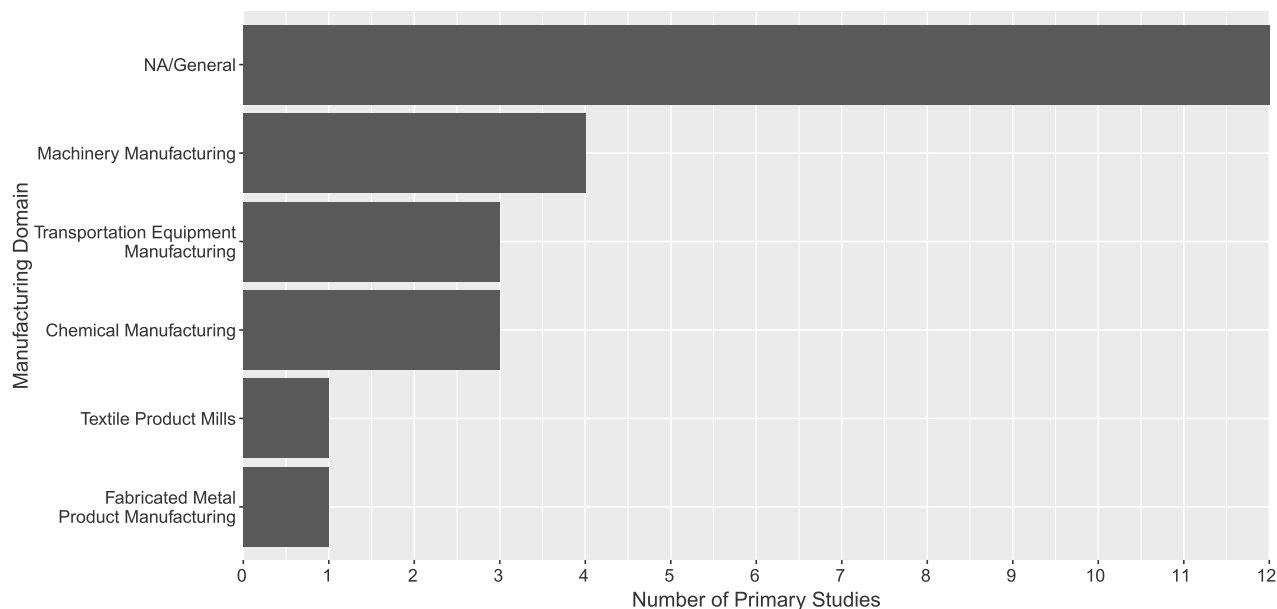


FIGURE 9. Knowledge graph manufacturing domains.

TABLE 11. Knowledge graph use cases.

Use Case	Studies
Knowledge Fusion	[S1] [S2] [S5] [S6] [S7] [S13] [S16] [S17] [S19] [S20] [S22] [S23]
Creation of Digital Twin	[S3] [S4] [S15]
Automated Process Integration	[S10] [S14] [S18]
Program Generation	[S8] [S11] [S21]
Predictive Analytics	[S12]
Knowledge Graph Platform	[S9]
No Information Provided	[S24]

of a data visualization and analysis system. In 2 cases a digital twin has been created, 2 studies supported the modeling/description of a cyber-physical system, and 2 studies provided a simulation platform. In 2 studies a system for the automated construction of a knowledge graph was proposed, 1 study proposed a KG management system. In 2 primary studies, no concrete system based on a knowledge graph was developed. Table 12 lists which kinds of system have been described by each primary study.

V. DISCUSSION

In this section, we provide a discussion of the summarized results obtained from the classification and analysis of our study, along with an identification of gaps and opportunities for future research, and a discussion of the threats to validity of our study.

TABLE 12. Knowledge graph-based system kinds.

Use Case	Studies
Search-based Application	[S1] [S4] [S7] [S12] [S13] [S14] [S19] [S22]
Code Generation System	[S8] [S11] [S21]
Digital Twin	[S3] [S15]
Data Visualization and Analysis System	[S2] [S17]
CPS Modelling Tool	[S5] [S6]
Simulation Platform	[S10] [S18]
Knowledge Graph Generator	[S20] [S23]
KG Management Platform	[S9]
No Information Provided	[S16] [S24]

A. ABOUT THE BIBLIOMETRIC KEY FACTS OF KG FOR PUBLICATIONS (RQ1)

The first aspect we want to highlight is the fact that the application of knowledge graphs in manufacturing and production (although in an early stage) is attracting a lot of attention from both industry and academia. It should be noted that most primary studies that have been considered in this literature review stem from the field of Computer Science. The Engineering and Business, Management, and Accounting sectors are far behind. This may be because the Computer Science community has led the way in knowledge representation models from the beginning. Today, knowledge graphs are understood as the natural evolution of such models to make them more adaptable to novel paradigms based on cloud and edge web platforms as well as new solutions for massive data and knowledge storage and maintenance. Furthermore, as can be observed in the results shown in the previous section, most of the studies selected have been published in conference proceedings to date. This is often an indicator that

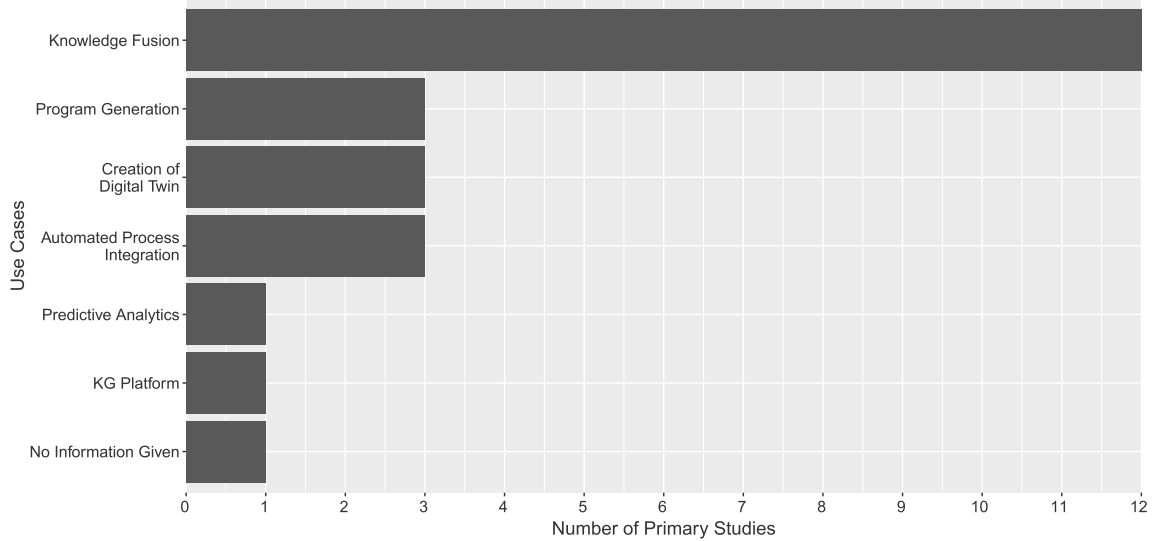


FIGURE 10. Knowledge graph use cases.

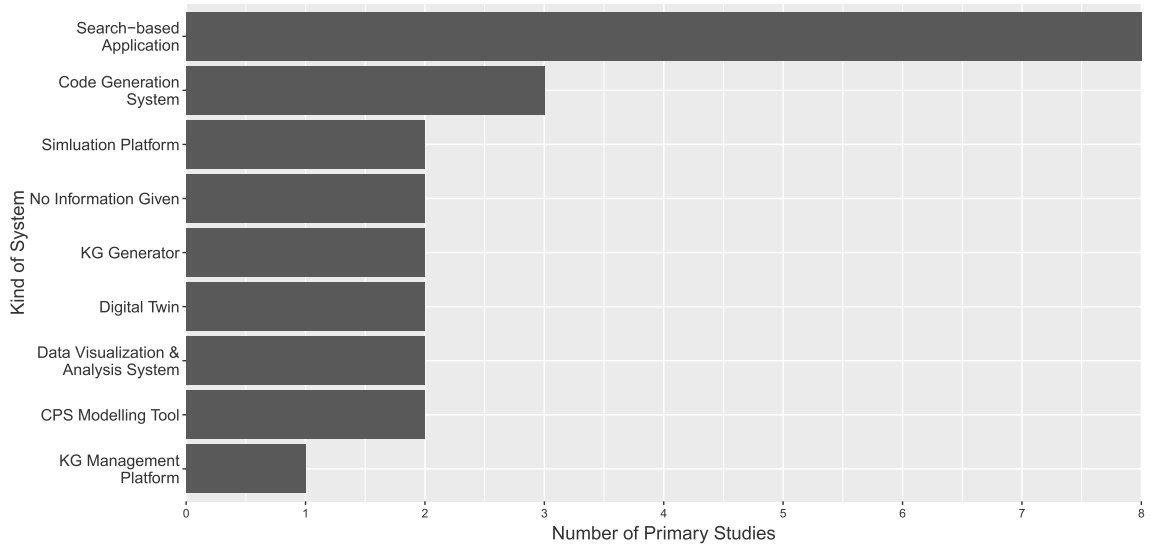


FIGURE 11. Knowledge graph system kinds.

the research topic is still emerging. In fact, in most research fields, as the state-of-the-art matures, more journal papers holding archival value are produced. Therefore, knowledge graphs do not seem to be an exception.

With regard to the countries leading research in knowledge graphs in the field of manufacturing and production, it could be observed that Germany and China are the two countries from which most primary studies originate from. Both countries also belong to the top manufacturing countries on a value-added basis, i.e., in 2018 China was the top manufacturing country and Germany held position 4.¹² Analyzing the relation between the origin countries of the primary studies

¹²See <https://www.statista.com/chart/20858/top-10-countries-by-share-of-global-manufacturing-output>

and the leading manufacturing countries further reveals that the United States as second-largest manufacturing nation only takes the fourth rank in the origin countries of the primary studies. For Japan as the third-largest manufacturing country, we could not find any primary study. Since the application of KGs in manufacturing and production is still an emerging topic (see above) and there are still little primary studies, the analysis of the origin countries should be seen as a current snapshot, which can easily change in the future.

B. ABOUT THE RESEARCH TYPE FACETS THAT THE IDENTIFIED PUBLICATIONS ADDRESS (RQ2)

It is also noteworthy that the number of validation research papers, which provide sound evidence of the usefulness of

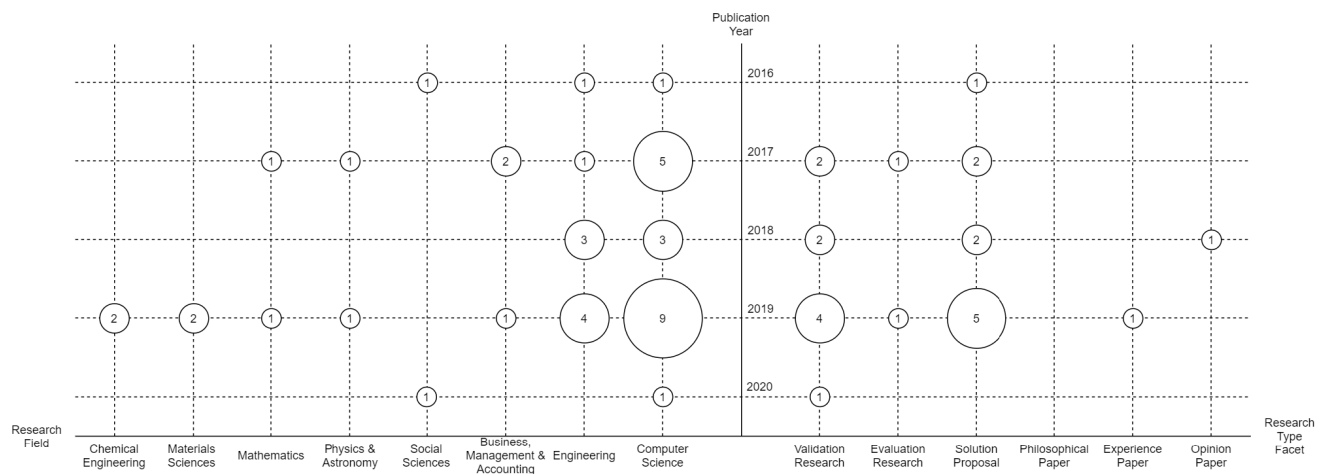


FIGURE 12. Summary of classification.

knowledge graphs in the production and manufacturing sector is still rather limited. So far most articles are proposing a solution for potential issues that can occur in production environments. However, these works have not been implemented in a real-world setting. Despite results obtained from validation research papers provide solid foundations for the employment of knowledge graphs in manufacturing and production, and show the usefulness of such approaches in this field. Knowledge graphs in the field of manufacturing and production is a very important path that cannot be ignored if we take a look at the evolution of topics, the number of publications, and research type forum over time as shown in Figure 12. At the same time, experience papers and opinion papers are in the minority at the moment. The fact that the new approaches based on deep learning have facilitated and accelerated the proposal of fully automated solutions does not seem to provide new views on their potential impact on production and manufacturing companies. This can probably change when advances are made to process numerical and tabular data in knowledge graphs, or on the other hand, to apply deep learning models directly on the knowledge graph without requiring an intermediate step of calculating embeddings¹³ from the graph, which can serve as direct input for the deep learning model.

C. ABOUT THE SPECIFIC KNOWLEDGE GRAPH CHARACTERISTICS (RQ3)

Another interesting takeaway is that against the vast majority of KG literature, knowledge graphs in an industry setting are usually constructed following a top-down approach. Thus, it seems that in an industry setting top-down knowledge graphs enjoy higher popularity. This could be caused by the fact that production settings require dedicated restrictions and schema constraints to ensure high quality in the knowledge graphs. With regard to the used graph model, RDF graphs

¹³Note: as mentioned in the Introduction, a KG embedding represents the data in the graph as numeric vectors.

have been used in a majority of the analyzed approaches. This is in contrast to the industrial state-of-practice that typically favors property graphs. However, the popularity of RDF graphs for KG development might hint at the greater potential of RDF for the development of knowledge-based applications.

It is worth mentioning that barely any paper with respect to this area is using machine learning or deep learning methods to expand the knowledge graph. We guess that this observation could be caused by the fact that companies want to know why decisions have been made and the missing explainability of 'black box' models such as neural networks is an uncertainty factor companies do not want to deal with in their production process. Additionally, as knowledge graphs mainly deal with textual pieces of information the number of machine learning tools that can be used for the prediction of categories is rather limited or have unrealistic assumptions to provide adequate results, at least if compared to situations that require the processing of numerical information, as is usually the case in the industrial sector.

In short, although the options in terms of representation and evolution of the knowledge graphs are diverse, most of the solutions proposed to date are eminently conservative and do not yet make intensive use of the new developments with regards to machine learning. We envision that, however, these advances will be increasingly incorporated into the existing body of literature soon as new solutions based on bottom-up knowledge graphs will be proposed in combination with more mature deep learning techniques that allow being used in still unexplored yet relevant domains.

D. ABOUT THE APPLICATION SCENARIOS (RQ4)

With regard to the application domain analysis, which answers where knowledge graphs have been employed in manufacturing and production, it could be observed that for 50 % of the primary studies no explicit application domain could be determined. We see two reasons for this: First, many

primary studies are still at an early stage of research (see also RQ2) and have not yet been evaluated in an industrial setting. Therefore, no concrete field of manufacturing could have been determined. Second, some primary studies propose general solutions that are not only applicable to a single field of manufacturing. For instance, analyzing data from production lines in a knowledge graph is a relevant topic across many manufacturing domains.

With respect to the use cases for which knowledge graphs have been applied, it could be observed that knowledge fusion is the main use case for knowledge graph applications, which has been exploited by half of the analyzed studies. This comes as no surprise, since the integration of different data sources can be seen as a strength of knowledge graphs. Further, knowledge fusion can also be seen as a foundation/prerequisite for other use cases, i.e., the creation of digital twins, automated process integration, program generation, and predictive analytics. All those use cases build upon the integration and analysis of data from different data sources. The spectrum of use cases also shows the potential of knowledge graphs as the underlying technology for production and manufacturing companies.

Analyzing the kinds of systems that have been developed in the primary studies revealed that in most cases (8 studies) search-based applications have been built based on knowledge graphs. These systems provide means for searching and exploring the data stored in a knowledge graph and allow analyzing relationships between data that has initially been stored in different data sources. Closely related to search-based applications are data visualization and analysis systems, which have been subject of research in 2 studies. All these systems focus on supporting analytical tasks and knowledge discovery. Code generation systems have been developed as part of 3 studies. In 2 studies, source code for autonomous robots/vehicles was generated, in 1 case test cases were generated from specifications. This shows the possibility of using knowledge graphs for automating constructive activities, i.e., software development tasks, which represents a paradigm shift from requirements-driven to data-driven development [71]. Multiple system kinds deal with digital representations manufacturing and production processes. This includes the creation of digital twins, the automated modeling (documentation) of cyber-physical systems, and the simulation of processes. Finally, 3 studies focus on the construction and management of knowledge graphs, rather than on use cases where knowledge graphs are further used. 2 of these approaches investigate the automated construction of knowledge graphs from existing data. 1 study has proposed a platform for the creation and management of knowledge graphs. These works can be seen as the foundation for facilitating the wider application of knowledge graphs in the future.

In summary, the analysis of the current application scenarios of knowledge graphs in manufacturing and production shows the wide applicability and potential of knowledge graphs as underlying technology and foundation for a variety

of different use cases and different system kinds. The different use cases (RQ4), the current number of primary studies (RQ1), and the current maturity of research (RQ3) show that knowledge graphs are still an emerging research topic with multiple open research challenges.

E. OPEN RESEARCH CHALLENGES

After examining the existing literature on knowledge graphs in the field of manufacturing and production, we have identified five open research challenges that have not been adequately addressed to date. The following is a description of each of these open research challenges (ORCs).

1) ORC1: PROPER HANDLING OF NUMERIC AND TABULAR DATA

Today, almost all of the solutions presented in the literature are specifically designed to work with the information of an eminently textual nature. While it is true that this is an important type of information in industry, it is not precisely the predominant kind in manufacturing and production environments that usually work with machines and equipment that produce data of a numerical and even tabular nature. Tabular data is the type of data that is frequently represented in comma-separated values and is usually one of the most common input methods in industrial environments since it allows to model a wide variety of data associated with temporal aspects (timestamps), spatial aspects (coordinates), and similar categorical attributes. However, such files usually lack extensive metadata, i.e., there is often no contextual information to the data. For interpreting such plain textual files (e.g., comma-separated values or Excel spreadsheets), knowledge graphs can open a wide range of possibilities for downstream tasks like data analysis, data integration or knowledge discovery. Recently, intensive research work has been carried out on the problem of mapping tabular data to knowledge graphs¹⁴ and it is likely that these advances may represent a qualitative leap forward in this sector.

2) ORC2: FURTHER RESEARCH IN REAL-TIME KNOWLEDGE GRAPHS

A fact that seems widely assumed among researchers and practitioners is the suitability of knowledge graphs to properly deal with data, information, and knowledge of different nature that might arrive through different channels and sources. However, in the course of this study, we have not observed a great number of works related to the temporal aspect when processing knowledge graphs. The manufacturing industry, however, requires solutions that can operate satisfactorily in environments with significant time constraints as many of the processes are automated and require a high degree of synchronization between them. In fact, we are convinced that further research on issues related to the fast updating knowledge graphs in real-time as information is received through different channels and sources will be a great advance

¹⁴<https://www.cs.ox.ac.uk/isg/challenges/sem-tab/>

in this sector. We look forward to new proposals in this regard when basic research would facilitate it.

3) ORC3: AUTOMATIC LINKAGE WITH OTHER PUBLICLY AVAILABLE KNOWLEDGE GRAPHS

One of the main characteristics that have made knowledge graphs popular is the possibility to grow based on identifying similar pivot points in other knowledge graphs that have been designed independently but have ended up being offered to the public, for example, on the web. However, it seems that all the research that has been published in relation to the manufacturing and production sector do not take into account this characteristic when designing and implementing solutions. On the contrary, it seems that the community is working with knowledge graphs contained in silos that solve a specific task without requiring connection to other publicly available graphs. This means that not all the capabilities that knowledge graphs can offer are being used to the fullest. Obviously, to stimulate more intensive research in this direction, it is necessary to have better solutions for the automatic alignment of entities between different knowledge graphs. In this sense, it should be noted that promising methods have appeared recently¹⁵ and will soon be used with certain guarantees of success.

4) ORC4: A HIGHER DEGREE OF COVERAGE OF ALL MANUFACTURING AND PRODUCTION DOMAINS

Based on the current state of evaluation (RQ2), although we have seen a great deal of research work in a number of categories in the manufacturing and production sector, the truth is that there is still a lack of validation of existing approaches in this context as well as many industrial categories that remain largely unexplored. For example, there is an abundance of generic solutions for Industry 4.0 or the Internet of Things, and even there are already some innovative solutions for categories that are important from the economic point of view such as the chemical and automotive sectors. However, we can see that most categories of the NAICS classification have not yet been explored from the point of view of basic and applied research on knowledge graphs to automate and/or improve the industrial processes that they currently implement (evaluation research). We think that this question is related to the incipient emergence of knowledge graphs in this sector. Therefore, it is expected that in the coming years more evaluation research will appear in each category that make up the NAICS taxonomy.

5) ORC5: COLLECTION OF BEST PRACTICES CONCERNING THE DEVELOPMENT OF REFERENCE ARCHITECTURES FOR KG DEVELOPMENT

It also appears from our study that no work has been done in the direction of compiling and understanding what the best practices are in this sector. It remains a great future challenge to be able to identify the reference architectures for the design, implementation, and exploitation of knowledge

graphs in industrial and production environments. We think that a compilation of best practices in this area can be truly beneficial in maintaining high standards of quality results, as well as, saving resources in the form of money and time when developing new systems or making changes to existing ones. The truth is that until now all the development has been carried out without clear guidelines that might promote the best outcomes. On the contrary, each team has developed its own solutions to the best of its knowledge. We believe that it could be highly beneficial to have a set of best practices that can standardize the development of knowledge graphs in the field of manufacturing and production. In this way, the domain can grow sustainably and reach a high degree of development that is beneficial to all stakeholders involved.

F. THREATS TO VALIDITY

To mitigate threats to validity [72] regarding subjective measures, all selected primary studies were analyzed according to defined criteria by all three authors. Studies with conflicting votes were analyzed in detail and discussed in the group before they were re-evaluated again by all researchers.

The exclusion of relevant studies during screening is another threat to validity. We tried to mitigate this by clearly defined criteria and a multi-stage screening process, which facilitated in-depth analysis of studies with ambiguous evaluations.

Our mitigation strategy against low statistical power is to use five academic libraries and Google Scholar as an additional data source to obtain the most complete possible result set of publications. We also considered alternative terms for describing sub-fields in the field of manufacturing and production as keywords for the search process to mitigate risks of publication selection and instrumentation.

VI. CONCLUDING REMARKS

In this paper, we have presented a systematic literature review on knowledge graphs in manufacturing and production environments. With the growing amount of scientific literature on KGs, an overview of the current development is needed to assess the applicability for this specific domain. Thus, we have come to weave a systematic literature review of one of the most prominent fields of application for knowledge graphs. To date, a lot of attention has been paid to fundamental research in KGs as attested by the growing literature on knowledge graph surveys.

After carrying out a systematic search for primary studies, we have classified the identified studies according to four facets: (1) bibliometric key facts, (2) research type, (3) knowledge graph characteristics, and (4) application scenarios. Based on this classification, we have analyzed the current state of research and identified open research challenges.

We identified that almost 90 % of the research has been conducted in one of the top 9 manufacturing countries worldwide.¹⁶ This highlights the fact that KGs are of particular importance for the manufacturing and production

¹⁵<https://paperswithcode.com/task/entity-alignment>

¹⁶See, Statista.

domain. Furthermore, the majority of research papers have been published in conferences. This fact illustrates that this research area is still in an early stage. It should also be noted that more than 95 % of all published studies underwent a review process and are thus academically sound. Another interesting fact is that almost all papers belong to the category of computer science or engineering. Barely any evaluation papers were identified and more than 80 % of the selected studies are either solution proposals or validation research papers. Notably, almost 70 % of all KGs in manufacturing and production follows a top-down approach whereas slightly more than 30 % are constructed employing a bottom-up approach. Finally, 87.5 % of the relevant KGs have been identified as RDF graphs whereas 12.5 % belong to the category of property graphs.

The general trend that can be seen is that this area of knowledge, albeit upward, is still very fragmented. This means that several research groups, as well as public and private organizations, focus their research on small sections and manufacturing sectors that are of interest to them, but they do not usually look further beyond. In addition, most solutions are focused on the classical use of graph-oriented computing, so it does not take full advantage of new methods based on deep learning that allows the processing of graphs. Nevertheless, although timidly, some works are emerging and our forecast is that the number of works in this sense will shoot up in the near future. Furthermore, the research activity has not yet covered a good part of the categories included in the NAICS classification, but we believe that new works will appear that will cover practically all the categories.

Although an upward trend can be discerned, we think that the use of knowledge graphs in manufacturing industries has not yet fully reached its peak. But the figures evidence that is currently and most likely will remain a very active field of research in the near future. After all, knowledge graphs offer an effective and efficient way to address some of the problems that have traditionally plagued the manufacturing and production environments. So as future work, we will continue to monitor the literature in this sector to gain a better understanding of how knowledge graphs can boost innovation and ensure that there are no remaining gaps to fill.

APPENDIX. PRIMARY STUDIES

- [S1] S. Peroni, F. Vitali, Interfacing fast-fashion design industries with semantic web technologies: The case of imperial fashion, *Journal of Web Semantics* 44 (2017) 37–53.
- [S2] K. S. Aggour, V. S. Kumar, P. Cuddihy, J. W. Williams, V. Gupta, L. Dial, T. Hanlon, J. Gambone, J. Vinciguerra, Federated multimodal big data storage & analytics platform for additive manufacturing, in: 2019 IEEE International Conference on Big Data (Big Data), IEEE, 2019, pp. 1729–1738.
- [S3] M. Ringsquandl, S. Lamparter, R. Lepratti, P. Kröger, Knowledge fusion of manufacturing operations data using representation learning, in: IFIP International Conference on Advances in Production Management Systems, Springer, 2017, pp. 302–310.
- [S4] A. Banerjee, R. Dalal, S. Mittal, K. P. Joshi, Generating digital twin models using knowledge graphs for industrial production lines, in: Proceedings of the 2017 ACM on Web Science Conference, 2017, pp. 425–430.

- [S5] I. Grangel-González, L. Halilaj, M.-E. Vidal, S. Lohmann, S. Auer, A. W. Müller, Seamless integration of cyber-physical systems in knowledge graphs, in: Proceedings of the 33rd Annual ACM Symposium on Applied Computing, 2018, pp. 2000–2003.
- [S6] I. Grangel-González, L. Halilaj, M.-E. Vidal, O. Rana, S. Lohmann, S. Auer, A. W. Müller, Knowledge graphs for semantically integrating cyber-physical systems, in: International Conference on Database and Expert Systems Applications, Springer, 2018, pp. 184–199.
- [S7] L. He, P. Jiang, Manufacturing knowledge graph: A connectivism to answer production problems query with knowledge reuse, *IEEE Access* 7 (2019) 101231–101244.
- [S8] A. Kattapur, Roboplanner: autonomous robotic action planning via knowledge graph queries, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 2019, pp. 953–956.
- [S9] P. Haase, D. M. Herzig, A. Kozlov, A. Nikolov, J. Trame, metaphactory: A platform for knowledge graph management, *Semantic Web* 10 (6) (2019) 1109–1125.
- [S10] A. Eibeck, M. Q. Lim, M. Kraft, J-park simulator: An ontology-based platform for cross-domain scenarios in process industry, *Computers & Chemical Engineering* 131 (2019) 106586.
- [S11] A. Nayak, V. Kesri, R. K. Dubey, Knowledge graph based automated generation of test cases in software engineering, in: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, 2020, pp. 289–295.
- [S12] W. Duan, Y.-Y. Chiang, Building knowledge graph from public data for predictive analysis: a case study on predicting technology future in space and time, in: Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, 2016, pp. 7–13.
- [S13] X. Zhang, X. Liu, X. Li, D. Pan, Mmkg: an approach to generate metallic materials knowledge graph based on dbpedia and wikipedia, *Computer Physics Communications* 211 (2017) 98–112.
- [S14] Y. Zhao, Q. Liu, W. Xu, Open industrial knowledge graph development for intelligent manufacturing service matchmaking, in: 2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), IEEE, 2017, pp. 194–198.
- [S15] M. Ringsquandl, E. Kharlamov, D. Stepanova, S. Lamparter, R. Lepratti, I. Horrocks, P. Kröger, On event-driven knowledge graph completion in digital factories, in: 2017 IEEE International Conference on Big Data (Big Data), IEEE, 2017, pp. 1676–1681.
- [S16] F. Leijie, B. Yv, Z. Zhenyuan, Constructing a vertical knowledge graph for non-traditional machining industry, in: 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), IEEE, 2018, pp. 1–5.
- [S17] X. Li, S. Zhang, R. Huang, B. Huang, C. Xu, B. Kuang, Structured modeling of heterogeneous cam model based on process knowledge graph, *International Journal of Advanced Manufacturing Technology* 96 (9-12) (2018) 4173–4193.
- [S18] X. Zhou, M. Q. Lim, M. Kraft, A smart contract-based agent marketplace for the j-park simulator—a knowledge graph for the process industry.
- [S19] O. Tushkanova, V. Samoylov, Knowledge net: Model and system for accumulation, representation, and use of knowledge, *IFAC-PapersOnLine* 52 (13) (2019) 1150–1155.
- [S20] M. Zhao, H. Wang, J. Guo, D. Liu, C. Xie, Q. Liu, Z. Cheng, Construction of an industrial knowledge graph for unstructured chinese text learning, *Applied Sciences* 9 (13) (2019) 2720.
- [S21] R. Li, W. Dai, S. He, X. Chen, G. Yang, A knowledge graph framework for software-defined industrial cyber-physical systems, in: IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, Vol. 1, IEEE, 2019, pp. 2877–2882.
- [S22] U. Dombrowski, A. Reiswich, C. Imdahl, Knowledge graphs for an automated information provision in the factory planning, in: 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, 2019, pp. 1074–1078.
- [S23] T. Liebig, A. Maisenbacher, M. Opitz, J. R. Seyler, G. Sudra, J. Wissmann, Building a knowledge graph for products and solutions in the automation industry.
- [S24] K. Dimitrov, Leveraging knowledge graph embedding techniques for industry 4.0 use cases, *Cyber Defence in Industry 4.0 Systems and Related Logistics and IT Infrastructures* 51 (2018) 10.

REFERENCES

- [1] A. Blumauer and H. Nagy, *The Knowledge Graph Cookbook*, vol. 27. Vienna, Austria: Mono/Monochrom, 2020.

- [2] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia," *Semantic Web*, vol. 6, no. 2, pp. 167–195, 2015.
- [3] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, 2008, pp. 1247–1250.
- [4] M. K. Bergman, *Knowledge Representation Practionary*. Springer, 2018.
- [5] A. Carlson, J. Betteridge, R. C. Wang, E. R. Hruschka, and T. M. Mitchell, "Coupled semi-supervised learning for information extraction," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, 2010, pp. 101–110.
- [6] N. Nakashole, M. Theobald, and G. Weikum, "Scalable knowledge harvesting with high precision and high recall," in *Proc. 4th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2011, pp. 227–236.
- [7] D. Vrandečić and M. Krötzsch, "Wikidata: A free collaborative knowledgebase," *Commun. ACM*, vol. 57, no. 10, pp. 78–85, Sep. 2014.
- [8] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A core of semantic knowledge," in *Proc. 16th Int. Conf. World Wide Web (WWW)*, New York, NY, USA, 2007, pp. 697–706.
- [9] D. B. Lenat, "CYC: A large-scale investment in knowledge infrastructure," *Commun. ACM*, vol. 38, no. 11, pp. 33–38, Nov. 1995.
- [10] A. Singhal, *Introducing the Knowledge Graph: Things, Not Strings*. May 2012. [Online]. Available: <https://googleblog.blogspot.co.at/2012/05/introducing-knowledge-graph-things-not.html>
- [11] D. Sullivan, *A Reintroduction to Our Knowledge Graph and Knowledge Panels*. 2020. [Online]. Available: <https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels>
- [12] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang, "Knowledge vault: A Web-scale approach to probabilistic knowledge fusion," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 601–610.
- [13] J. F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA, USA: Addison-Wesley, 1983.
- [14] J. F. Sowa, "Conceptual graphs as a universal knowledge representation," *Comput. Math. Appl.*, vol. 23, nos. 2–5, pp. 75–93, Jan. 1992.
- [15] R. R. Bakker, "Knowledge Graphs: Representation and structuring of scientific knowledge," Ph.D. dissertation, Univ. Twente, Enschede, The Netherlands, 1987.
- [16] H. J. Smit, "Consistency and robustness of knowledge graphs," Ph.D. dissertation, Univ. Twente, Enschede, The Netherlands, 1991.
- [17] P. H. D. Vries, "Representation of scientific texts in knowledge graphs," Ph.D. dissertation, Dept. Behav. Social Sci., Univ. Groningen, Groningen, The Netherlands, 1989.
- [18] L. Ehrlinger and W. Wöb, "Towards a definition of knowledge graphs," in *Proc. Joint Proc. Posters Demos Track 12th Int. Conf. Semantic Syst.*, vol. 1695. Aachen, Germany: Technical Univ. of Aachen, Sep. 2016, pp. 13–16.
- [19] I. Robinson, J. Webber, and E. Eifrem, *Graph Databases*, 2nd ed. Newton, MA, USA: O'Reilly Media, 2015.
- [20] C. Feilmayr and W. Wöb, "An analysis of ontologies and their success factors for application to business," *Data Knowl. Eng.*, vol. 101, pp. 1–23, Jan. 2016.
- [21] D. Lukovnikov, A. Fischer, J. Lehmann, and S. Auer, "Neural network-based question answering over knowledge graphs on word and character level," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1211–1220.
- [22] Y. Zhang, H. Dai, Z. Kozareva, A. J. Smola, and L. Song, "Variational reasoning for question answering with knowledge graph," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.
- [23] J. Daiber, M. Jakob, C. Hokamp, and P. N. Mendes, "Improving efficiency and accuracy in multilingual entity extraction," in *Proc. 9th Int. Conf. Semantic Syst.*, 2013, pp. 121–124.
- [24] L. Dietz, A. Kotov, and E. Meij, "Utilizing knowledge graphs for text-centric information retrieval," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jun. 2018, pp. 1387–1390.
- [25] Z. Zheng, X. Si, F. Li, E. Y. Chang, and X. Zhu, "Entity disambiguation with freebase," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, vol. 1, Dec. 2012, pp. 82–89.
- [26] G. Zhu and C. A. Iglesias, "Exploiting semantic similarity for named entity disambiguation in knowledge graphs," *Expert Syst. Appl.*, vol. 101, pp. 8–24, Jul. 2018.
- [27] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on freebase from question-answer pairs," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1533–1544.
- [28] L. Heck and G. Tur, "Leveraging knowledge graphs for Web-scale unsupervised semantic parsing," in *Proc. INTERSPEECH*, 2013, pp. 1594–1598.
- [29] Z. Sun, J. Yang, J. Zhang, A. Bozzon, L.-K. Huang, and C. Xu, "Recurrent knowledge graph embedding for effective recommendation," in *Proc. 12th ACM Conf. Recommender Syst.*, Sep. 2018, pp. 297–305.
- [30] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "KGAT: Knowledge graph attention network for recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 950–958.
- [31] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowl.-Based Syst.*, vol. 151, pp. 78–94, Jul. 2018.
- [32] P. Ristoski and H. Paulheim, "RDF2vec: RDF graph embeddings for data mining," in *Proc. Int. Semantic Web Conf.* Berlin, Germany: Springer, 2016, pp. 498–514.
- [33] S. Auer, V. Kovtun, M. Prinz, A. Kasprzik, M. Stocker, and M. E. Vidal, "Towards a knowledge graph for science," in *Proc. 8th Int. Conf. Web Intell., Mining Semantics (WIMS)*, New York, NY, USA, 2018, pp. 1:1–1:6.
- [34] P. Ernst, C. Meng, A. Siu, and G. Weikum, "KnowLife: A knowledge graph for health and life sciences," in *Proc. IEEE 30th Int. Conf. Data Eng.*, Mar. 2014, pp. 1254–1257.
- [35] I. Abdelaziz, A. Fokoue, O. Hassanzadeh, P. Zhang, and M. Sadoghi, "Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions," *J. Web Semantics*, vol. 44, pp. 104–117, Mar. 2017.
- [36] C. Xie, P. Yang, and Y. Yang, "Open knowledge accessing method in IoT-based hospital information system for medical record enrichment," *IEEE Access*, vol. 6, pp. 15202–15211, 2018.
- [37] L. Li, P. Wang, J. Yan, Y. Wang, S. Li, J. Jiang, Z. Sun, B. Tang, T.-H. Chang, S. Wang, and Y. Liu, "Real-world data medical knowledge graph: Construction and applications," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101817.
- [38] M. Iannacone, S. Bohn, G. Nakamura, J. Gerth, K. Huffer, R. Bridges, E. Ferragut, and J. Goodall, "Developing an ontology for cyber security knowledge graphs," in *Proc. 10th Annu. Cyber Inf. Secur. Res. Conf.*, Apr. 2015, pp. 1–4.
- [39] C. Asamoah, L. Tao, K. Gai, and N. Jiang, "Powering filtration process of cyber security ecosystem using knowledge graph," in *Proc. IEEE 3rd Int. Conf. Cyber Secur. Cloud Comput. (CSCloud)*, Jun. 2016, pp. 240–246.
- [40] Z. Han, X. Li, H. Liu, Z. Xing, and Z. Feng, "DeepWeak: Reasoning common software weaknesses via knowledge graph embedding," in *Proc. IEEE 25th Int. Conf. Softw. Anal., Evol. Reeng. (SANER)*, Mar. 2018, pp. 456–466.
- [41] Y. Deng, D. Lu, D. Huang, C.-J. Chung, and F. Lin, "Knowledge graph based learning guidance for cybersecurity hands-on labs," in *Proc. ACM Conf. Global Comput. Educ.*, May 2019, pp. 194–200.
- [42] J. A. M. B. Josko, L. Ehrlinger, and W. Wöb, "Towards a knowledge graph for data defects," in *Proc. 11th Int. Conf. Adv. Databases, Knowl., Data Appl.* Athens, Greece: International Academy, Research Industry Association, Jun. 2019, pp. 57–60.
- [43] M. Sette, L. Tao, and N. Jiang, "A knowledge-driven Web tutoring system framework for adaptive and assessment-driven open-source learning," in *Proc. IEEE 41st Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, vol. 2, Jul. 2017, pp. 712–717.
- [44] Y. Chi, Y. Qin, R. Song, and H. Xu, "Knowledge graph in smart education: A case study of entrepreneurship scientific publication management," *Sustainability*, vol. 10, no. 4, p. 995, Mar. 2018.
- [45] F. Dang, J. Tang, and S. Li, "MOOC-KG: A MOOC knowledge graph for cross-platform online learning resources," in *Proc. IEEE 9th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC)*, Jul. 2019, pp. 1–8.
- [46] D. Fensel, U. Şimşek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich, and A. Wahler, *Knowledge Graphs: Methodology, Tools and Selected Use Cases*. Springer Nature, 2020.
- [47] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proc. IEEE*, vol. 104, no. 1, pp. 11–33, Jan. 2016.
- [48] J. Z. Pan, G. Vetere, J. M. Gomez-Perez, and H. Wu, *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer, 2017.
- [49] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic Web*, vol. 8, no. 3, pp. 489–508, Dec. 2016.

- [50] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017.
- [51] Y. Lin, X. Han, R. Xie, Z. Liu, and M. Sun, "Knowledge representation learning: A quantitative review," 2018, *arXiv:1812.10901*. [Online]. Available: <http://arxiv.org/abs/1812.10901>
- [52] J. Yan, C. Wang, W. Cheng, M. Gao, and A. Zhou, "A retrospective of knowledge graphs," *Frontiers Comput. Sci.*, vol. 12, no. 1, pp. 55–74, Feb. 2018.
- [53] G. A. Gesese, R. Biswas, and H. Sack, "A comprehensive survey of knowledge graph embeddings with literals: Techniques and applications," in *Proc. Workshop Deep Learn. Knowl. Graphs*, 2019, pp. 31–40.
- [54] S. M. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupard, "Representation learning for dynamic graphs: A survey," *J. Mach. Learn. Res.*, vol. 21, no. 70, pp. 1–73, 2020.
- [55] M. Kejrival, *Domain-Specific Knowledge Graph Construction*. Springer, 2019.
- [56] X. Wang and S. Yang, "A tutorial and survey on fault knowledge graph," in *Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health*. Springer, 2019, pp. 256–271.
- [57] T. Al-Moslmi, M. G. Ocana, A. L. Opdahl, and C. Veres, "Named entity extraction for knowledge graphs: A literature overview," *IEEE Access*, vol. 8, pp. 32862–32881, 2020.
- [58] N. Heist, S. Hertling, D. Ringler, and H. Paulheim, "Knowledge graphs on the Web—An overview," 2020, *arXiv:2003.00719*. [Online]. Available: <http://arxiv.org/abs/2003.00719>
- [59] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition and applications," 2020, *arXiv:2002.00388*. [Online]. Available: <http://arxiv.org/abs/2002.00388>
- [60] A. Hogan, E. Blomqvist, A.-C. N. Ngomo, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, R. Navigli, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, and A. Zimmermann, "Knowledge graphs," 2020, *arXiv:2003.02320*. [Online]. Available: <http://arxiv.org/abs/2003.02320>
- [61] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, Apr. 2007.
- [62] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *School Comput. Sci. Math., Keele Univ., Keele, U.K., Tech. Rep. EBSE-2007-01*, 2007.
- [63] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, "Requirements engineering paper classification and evaluation criteria: A proposal and a discussion," *Requirements Eng.*, vol. 11, no. 1, pp. 102–107, Mar. 2006.
- [64] V. Alves, N. Niu, C. Alves, and G. Valença, "Requirements engineering for software product lines: A systematic literature review," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 806–820, Aug. 2010.
- [65] M. Kaufmann, G. Wilke, E. Portmann, and K. Hinkelmann, "Combining bottom-up and top-down generation of interactive knowledge maps for enterprise search," in *Proc. Int. Conf. Knowl. Sci., Eng. Manage.* Springer, 2014, pp. 186–197.
- [66] Z. Zhao, S.-K. Han, and I.-M. So, "Architecture of knowledge graph construction techniques," *Int. J. Pure Appl. Math.*, vol. 118, no. 19, pp. 1869–1883, 2018.
- [67] S. Das, J. Srinivasan, M. Perry, E. I. Chong, and J. Banerjee, "A tale of two graphs: Property graphs as RDF in oracle," in *Proc. EDBT*, 2014, pp. 762–773.
- [68] O. Hartig, "Reconciliation of RDF* and property graphs," 2014, *arXiv:1409.3288*. [Online]. Available: <http://arxiv.org/abs/1409.3288>
- [69] D. Alocci, J. Mariethoz, O. Horlacher, J. T. Bolleman, M. P. Campbell, and F. Lisacek, "Property graph vs RDF triple store: A comparison on glycan substructure search," *PLoS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0144578.
- [70] R. Angles, H. Thakkar, and D. Tomaszuk, "RDF and property graphs interoperability: Status and issues," in *Proc. 13th Alberto Mendelzon Int. Workshop Found. Data Manage.*, Asunción, Paraguay, 2019, pp. 1–11.
- [71] J. Bosch, H. H. Olsson, and I. Crnkovic, "It takes three to tango: Requirement, outcome/data, and AI driven development," in *Proc. SiBW*, 2018, pp. 177–192.
- [72] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer, 2012.



GEORG BUCHGEHER received the Ph.D. degree in 2013. He has studied business and computer science at the Johannes Kepler University Linz. Since 2006, he has been working with SCCH. He is currently a Key Researcher and a Project Manager in software science with the Software Competence Center Hagenberg GmbH (SCCH). He is also an External Lecturer with the University of Applied Sciences Upper Austria, Campus Wels. He has over 15 years of experience in fundamental and

applied research projects. He has published over 35 scientific publications at international conferences and in scientific journals. His research interests include software architectures, software engineering, component- and service-based development, microservices, and knowledge graphs. He is a program committee member for multiple international conferences and workshops and acts as a reviewer for international journals.



DAVID GABAUER received the B.Sc. and M.Sc. degrees in economics from Johannes Kepler University (JKU), Austria, and the Ph.D. degree in social and economic sciences from the Vienna University of Economics and Business, Austria. He is currently a Senior Data Scientist in data science with the Software Competence Center Hagenberg (SCCH), Austria, where he is involved in several applied and fundamental industrial research projects related to quality control and

predictive maintenance. He has published more than 25 articles in widely renowned academic journals. His research interests include machine learning, deep learning, computer vision, pattern recognition, international finance, and time series econometrics.



JORGE MARTINEZ-GIL received the Ph.D. degree in computer science from the University of Malaga, in 2010. He has held a number of research position across some European countries, such as Austria, Germany, and Spain. He currently holds a research position in data science with the Software Competence Center Hagenberg, Austria, where he is involved in several applied and fundamental research projects related to knowledge-based technologies. He is a Spanish-born Computer Scientist

working in data and knowledges engineering. He is also involved in the evaluation of many scientific manuscripts, project proposals, and study programs in ICT related fields.



LISA EHRLINGER received the Diploma degree in computer science from Johannes Kepler University (JKU), where she is currently pursuing the Ph.D. degree in automated continuous data quality measurement. She is also a Senior Researcher in data science with SCCH and JKU Linz, Austria. At JKU, she holds exercises and lecture blocks about information systems and ontology engineering. At SCCH, she is a Project Manager in the multi-firm COMET-project Sebista (Secure Big Stream Data Processing) with company partners, such as voestalpine Stahl GmbH and KTM Innovation GmbH. Her research interests and publications cover the topics data quality, knowledge graphs, NoSQL databases, and information integration. She was and is a program committee member at several scientific workshops, such as GraphSM, MLKgraphs, and QEKGraph, and the DBKDA conference. She was invited for a talk at the premier MIT Chief Data Officer and Information Quality Symposium to present her most recent research on data quality tools, in 2019 and 2020, respectively.

• • •