SEVEN
publicações acadêmicas

Knowledge graph: A strategy for knowledge management?

https://doi.org/10.56238/sevened2023.006-041

Valéria Macedo
(IBICT/UFRJ)

Larriza Thurler
(IBICT/UFRJ)

Elaine Dias
(IBICT/UFRJ)

Marcos Cavalcanti
(UFRJ)

ABSTRACT
This exploratory study sought to deepen the understanding of the benefits of using the knowledge graph technological solution for knowledge management. Through a bibliometric analysis of 45 academic articles selected from the Scopus database, with no time range determined, it was observed that the term has been addressed in academic productions on various topics with a greater number of publications from Asian researchers (71%). To understand the relationships of the keywords used by the authors of these articles, the VOSviewer tool was used to create the graph, making it possible to visualize the terms' clusters. It was identified that the theme is emerging, as the first article published in the Scopus database took place in 2011 and an expressive growth of publications can be observed in 2020. The production of academic knowledge with the use of the coupled knowledge graph stands out to the design of algorithms that support tutoring and curation activities with a focus on knowledge management in the areas of Health - including during the Covid-19 pandemic - and Education. The transversality of this solution is also noted, with studies using knowledge graphs in the areas of Engineering and Computer Science being observed. After further analysis, the authors selected 12 articles that presented initiatives focused on the area of Management and the use of the technological solution in activities aimed at intelligent recommendation can be observed, including the adoption of artificial intelligence and sampling algorithms for the study and development of solutions that resulted in the visualization of knowledge bases through the knowledge graph. The trend of using knowledge graphs for knowledge management pointed out by Knowledge Management (KM) World magazine can be found in business environments, such as Google and Wikipedia. It is believed that more empirical studies with knowledge graphs for knowledge management will be intensified in the digital age and this can be followed up in scientific congresses on the subject.

Keywords: Knowledge graph, Knowledge management, Knowledge database.

1 INTRODUCTION

Every year, the Knowledge Management (KM) World magazine[1] lists the main trends in knowledge management (KM) and for 2021 the highlights were the technological solutions that will be of great value to organizational strategies. In addition to technological solutions that enable digital

[1] KMWorld is recognized worldwide as one of the leading news outlets, study providers and conference organizers on knowledge management, document management and content management.

environments for collaborative work, the use of cloud technology, artificial intelligence and language processing solutions, *knowledge graphs* (KG) stand out.

It is possible to say that these solutions were enhanced with the advent of the pandemic and should contribute to the development and leverage of KM in organizations, demonstrating that the acceleration of digital transformation in the business world impacts changes in the use of KM strategies and practices in organizations in the coming years.

Collaborative remote work has had a big boost with the significant increase in companies providing video conferencing platform solutions. In this sector, new companies have emerged, with significant investments in technology due to the need for organizations and people to interact through online platforms. Even companies such as Google have offered educational solutions ("Google Classroom") free of charge to public schools as a tool option for use by teachers and students, continuing learning since the emergence of the pandemic.

Among the trends, the management of corporate content that involves records, processes, metadata and intelligent content capture, with the transfer of processes to the digital environment will require a significant increase *in datacenters* with the storage of data and information in the cloud for an intelligent management of the organization's information. And yet, according to KMWorld magazine, the adoption of artificial intelligence (AI) by various sectors of the economy, incorporating it into production processes, human resources management, marketing, sales, and service areas demonstrates a significant growth demanding professionals specialized in building bridges between humans and AI (BROWN et al, 2020).

In addition, text analysis and natural language processing (NLP) associated with AI benefit from the analysis of large volumes of information due to the ability to identify patterns, including relating and linking the knowledge explained in academic articles. And, according to KMWorld, this scenario was highlighted with the need for constant information about COVID-19.

The development and potential growth of KG are now considered in this emerging technology due to the speed of data processing, storage capacity, and the detection of relationships between data, favoring complex queries, enhancing the knowledge manager in the view of the leadership. The banking, healthcare, retail, e-commerce and transportation sectors have already adopted this tool to support the decision-making process. According to research conducted by Gartner, knowledge graphs enable data and analytics leaders to derive business value - enriching data with semantics, making it more intuitive and easy to interpret, favoring analysis and decision-making in leadership management.

In this scenario, the question arises as to what is the global academic production on the *knowledge graph* for knowledge management? This exploratory study seeks to understand the use of KG and the impacts of technological trends with a focus on the *knowledge graph* for knowledge management.

2 LITERATURE REVIEW

2.1 KNOWLEDGE GRAPH (KG)

Conceptually, there are a number of definitions, both in the academic and business environments, to explain what KG is and what it is for. For this study, the following definition is adopted: "the knowledge graph represents a collection of interconnected descriptions of entities - objects, events or concepts. These graphs incorporate data through links and semantic metadata, promoting a network that integrates, unifies, analyzes, and shares data in a contextualized way." [2]

KGs have three components: databases with structured data that make it easier to query; the knowledge base with a formal semantic structure so that the data can be interpreted and the knowledge graphs that unite the graph as a form of networked data analysis and its representativeness highlighted in the data framework.

According to Suchanek and Weikum (2013), large knowledge bases have been automated and enhanced with the emergence of a community for sharing data and information. Wikipedia is an example of these communities that were formed by extracting information from sources available on the web in a scalable way. To the Google Knowledge Graph, with the creation of a search system that results in knowledge cards, and to IBM's Watson initiative, with an automated system of answers and questions, are added the successful experiences of technologies that involve the use of big data, natural language, machine learning and artificial intelligence.

The purpose of this study is not to deepen the understanding of the techniques used for the processes that involve mining and extraction of knowledge from web pages, but to understand that there are important procedures that must be followed for the treatment and reading of knowledge bases, such as the creation of entities, semantic classes and relationships. A classic example of these knowledge bases is the use of intelligent computer behavior when presenting answers to questions and semantic questions that result in precise and concise answers, such as the virtual assistant of Banco Bradesco "Bia" or Magazine Luiza "Magalu".

According to Pan (2017), information systems use three components: construction; storage and consumption; and appropriate technologies for the creation of KGs. The areas of knowledge representation, databases, ontology and semantic web are used to deal with: (a) language, schemas and standard vocabularies for the representation of knowledge and use of reasoning; (b) databases, graphs and repositories for the storage of knowledge; (c) methodologies, publishing, and design standards to build knowledge engineering. Frameworks, mining technology, natural language processing, and machine learning are used for knowledge learning.

[2] The knowledge graph represents a collection of interlinked descriptions of entities – objects, events or concepts. Knowledge graphs put data in context via linking and semantic metadata and this way provide a framework for data integration, unification, analytics and sharing. Disponível em <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph> Acesso em 01/07/2021.

Moschitti et al (2017) believe that retrieving these answers from web users' questions in natural language is one of the most promising methods for organizations to evolve in the digital age. And KGs become fundamental tools for data processing, structuring the knowledge extracted from knowledge bases. In addition, the classic methods of organization, location and selection of information are research and development assets of companies, which, by incorporating relevant methodologies, technologies and systems, help to obtain relevant information for decision making.

> In enterprise information access scenarios, quality control systems are intended to replace traditional keyword-based search interfaces and allow users to more effectively exploit the structured knowledge in which the company has invested. This is likely the case for non-enterprise scenarios, such as web search, as well. However, there are some important differences between corporate and non-corporate knowledge graphs (Moschitti et al, 2017, p.193).[3]

It is possible to list three main characteristics of corporate knowledge graphs such as KG, according to Moschitti et al (2017):

a) Superior quality control due to centralized processes, compliance with consistency criteria, and because they are based on ontology, resulting in homogeneity in process development. Non-enterprise knowledge graphs have undisciplined processes with data structure and results that are difficult to understand.

b) Low diversity is a natural consequence of the fact that a company or organization is primarily interested in modeling knowledge about its business activities and the environment, rather than the entire world.

c) The structure and content of the questions in the corporate environment are more focused on the business domain than those carried out in the web environment, with organizations developing quality control systems from the bottom up, that is, they start with simple questions to prove the operation and then migrate to more complex approaches. And they also incorporate user feedback mechanisms in order to try to structure a more accurate system over time.

2.2 KNOWLEDGE MANAGEMENT (KM)

Ivanova and other researchers (2021) sought to analyze how much knowledge there is in KG and made some findings based on a theoretical review and empirical analysis of three case studies from the perspective of knowledge management. Initially, they revived how crucial knowledge is for organizations to create value through their tangible and intangible assets. However, they pointed out

[3] In enterprise information access scenarios, QA systems are aimed to replace traditional keyword-based search interfaces and allow users to exploit more effectively the structured knowledge in whose creation the enterprise has invested. This is probably the case for non-enterprise scenarios, such as Web search, as well. There are however some important differences between enterprise knowledge graphs and non-enterprise. (Moschitti et al, 2017, p.193)

that knowledge is not historically contextualized in a catalog, but incorporated into organizational routines and culture, in addition to the bonds between their employees. One of the main goals of organizations when they strategically manage their knowledge by connecting individuals was also highlighted, which is to promote the knowledge spiral by transforming tacit knowledge into explicit knowledge.

With technological evolution, can it be seen that knowledge bases have become an asset that stores collective knowledge? According to the authors, this is the new scenario in KM due to the various methods and processes adopted in the past when there was a need for human action. When analyzing publications presented at the International Web Conference (ISWC) in the years 2017 to 2019, it was found that KGs collaborate with organizational knowledge and make implicit knowledge explicit, generating knowledge creation when they become sources of knowledge represented by relational databases and unstructured documents organized in a KG. As a consequence, KG improves search functionality, i.e., the restructuring, integration and enrichment of data, providing knowledge generation.

3 RESEARCH METHOD AND MATERIALS

This study has an exploratory objective by conducting a bibliographic research and presents an approach of quantitative and qualitative analysis of the data and information captured. According to Severino (2007), bibliographic research focuses on obtaining available records - in databases such as published articles - and, through data and patterns already worked by other researchers, analyzing the researched themes.

The bibliometric study helps to understand how academia has dealt with the theme and disseminated it through impact indicators and also contributes to verifying the trend and growth of research interest in the area. For this study, some techniques of bibliometric study were used to understand the origin of the generation of knowledge on the subject and its evolution over time.

The use of the tools for data visualization helps to understand the relationships between the themes and sub-themes in the context of the research, with the possibility of analyzing the contexts and approaches of the studies with the formation of clusters, that is, the main nodes formed through the database.

The quantitative results explain the findings of this study, in which a database of articles captured in the Scopus database was used, on May 31, 2021, and a file with csv extension was generated. Using Excel, the data were structured and analyzed, and then submitted to the VOSViewer software (VAN ECK; WALTMAN, 2010), and then the qualitative analysis and the conception of the graphs were carried out.

3.1 USE FROM SCOPUS BASE

The bibliometric exploratory search was carried out in the Scopus database and used the following search terms: "knowledge graph" and "knowledge management" which mean "knowledge graph" and "knowledge management", and presented the following search syntax: ((KEY("knowledge management") AND KEY("knowledge graph"))) refined by Document types: (ARTICLE).

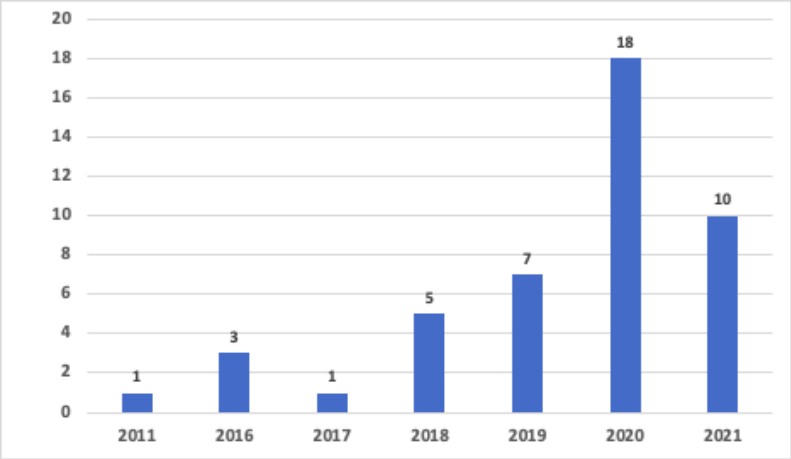
The research carried out on May 31, 2021 resulted in 45 scientific articles, without specifying the period for research. It is worth noting that there is evidence of several papers presented at conferences in various areas of knowledge, however this study prioritized only the search for scientific articles because they are peer-reviewed. Considering that the volume was not high, it was decided to leave it without restriction of time limit or other filter to refine the search, in addition to the definition of the type of document: article. It is observed that most of the articles are related to the areas of Computer Science (41), Engineering and Materials (26), Mathematics (6), Management (3), Social Sciences (3), considering that the same article appears in more than one category.

We chose to use the Scopus database because it is a multidisciplinary database that gathers abstracts and citations of peer-reviewed literature (such as scientific journals, books, congress processes and industry publications). It is one of the largest databases of scientific articles in the world, with more than 75 million records, more than 24,000 active titles (of these, 23,500 peer-reviewed journals) and 5,000 publishers. In addition, I think that the justification may be due to the free access to the Scopus database by the researchers the research was carried out on this basis and suggest in the conclusion that new bases, including papers presented at congresses, should be analyzed to understand the evolution of the theme in academia for knowledge management,

4 ANALYSIS OF RESULTS

The research resulted in a total of 45 scientific articles published during 10 years - May 2011 and May 2021 - related to KG and KM in the various areas of knowledge. The result is shown in Figure 1.

Figure 1: Publication by year on the knowledge graph (in absolute quantity) prepared by the authors.



Source: Scopus Base
Notes: Partial data for the year 2021; There was no publication in the period from 2012 to 2015.

It is observed that from 2018 onwards there was greater interest in the topic, with a peak of 18 articles published in 2020. This result may indicate greater interest in the topic in academia. In relation to the country of origin of the articles, there is a great representation and leadership of academic production in Asia, especially in China, as shown in chart 1 detailed below:

Table 1: Publication by country of origin.

Country of Origin	**Total**	**%**
Asia: China (28); Japan (1); South Korea (1); Croatia (1); Taiwan (1)	32	71%
North America: USA (7)	7	16%
European Community: France (1); Germany (1); Denmark (1); Spain (1); Greece (1)	5	11%
Latin America: Ecuador (1)	1	2%
Total Articles	45	100%

Source: Scopus Base
Notes: Partial data for the year 2021; There was no publication in the period from 2012 to 2015.

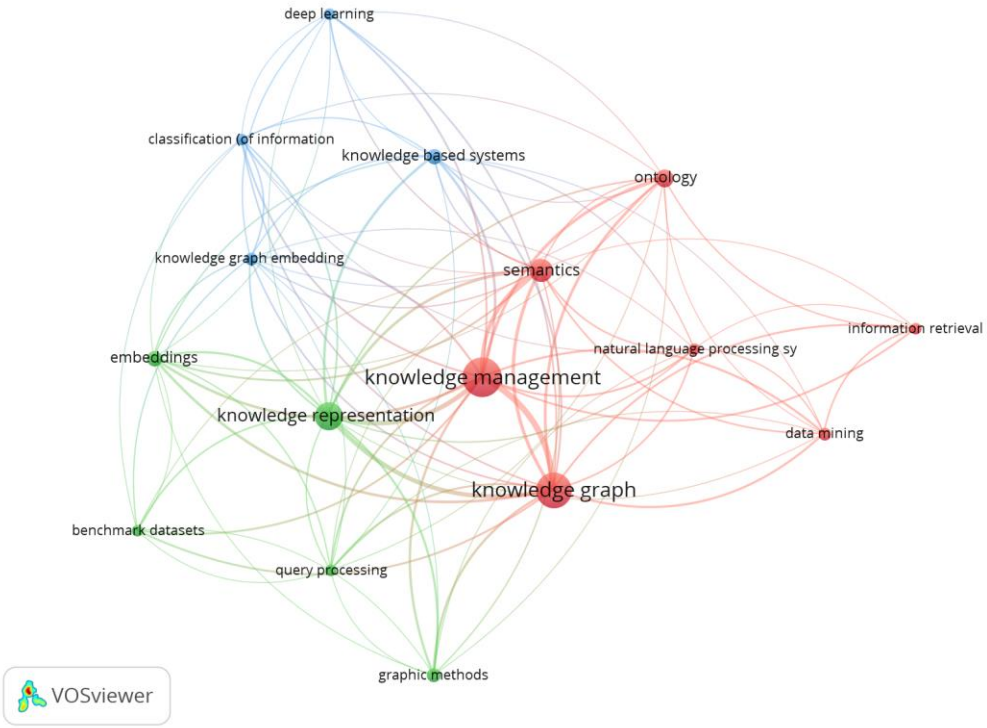
It should be noted that most of the studies were supported by foundations, research institutes and public agencies such as ministries of science and technology. It is noteworthy that most of the Chinese studies have obtained support and resources from the National Natural Science Foundation of China, demonstrating the importance of the state in supporting scientific research in the country.

There is also a diversity of journals that have published articles on the subject in various areas of knowledge, however, 3 (three) journals stand out: IEEE Access (7) which has a multidisciplinary approach; Journal of Information Science (3) and Journal of Computer Science and Technology, both focused on information science and technology, with more articles on the topics. Most journals with

only one (1) publication are focused on the fields of Engineering and/or Computer Science. All articles use the English language for publication, as expected because it is an international database.

In order to understand the relationships and *clusters* formed by the keywords used by the authors in the article, the following graph (Figure 2) was created using the VOSviewer tool:

Figure 2: Graph of the Clusters of keywords highlighted by the authors of the articles published in the Scopus Database. Own elaboration.



In the graph above, the theme of knowledge management *can be visually observed* as the central point of the articles analyzed, concluding that three *clusters* were formed highlighted in red, green and blue. In red, the *cluster* grouped the following topics: knowledge management, KG, and topics related to semantics and areas of application of ontologies (such as information retrieval, data mining, and natural language processing systems). In the green cluster, the highlight was the representation of knowledge and themes such as word associations, graphic methods, databases and data processing in these databases. And finally, in the blue cluster, the approach was with topics such as *deep learning*, information classification and other technologies to process a large volume of data.

Chart 2 shows an analysis of the context of 12 articles chosen, after reading the abstracts available by the authors of this study, to understand the use of KG in academic studies:

Table 2: Analysis of the context of 12 selected articlesSource: prepared by the authors.

Article Title	Authors	Context
1. Applying graph sampling methods on student model initialization in intelligent tutoring systems. (Croácia)	VIŠTICA, Marija; GRUBIŠIC, Ani; RYE, Branko. (2016)	Comparative study of graphical sampling algorithms (Random Walk (Deo and Gupta, 2001), Metropolis-Hastings Random Walk (Metropolis, Rosenbluth, Rosenbluth, Teller, Teller, 1953), (Hastings, 1970)), Forest Fire (Leskovec, Kleinberg and Faloutsos, 2005), Snowball ((Coleman, 1958), (Goodman, 1961)) and Represent algorithm (Grubišić, 2012) collaborated in the modeling of an intelligent tutoring system for the student, that is, through the use of these algorithms a sample of knowledge is generated that truly represents a particular field of knowledge.
2. Modeling the Correlations of Relations for Knowledge Graph Embedding. (China)	ZHU, Ji-Zhao et al. (2018)	Proposal of a method to better correlate the entities and relationships in a KG.
3. Structural Modeling of Heterogeneous CAM Model Based on Process Knowledge Graph. (China)	LI, Xiuling et al. (2018)	Proposal of a KG on the CAM (computer aided manufacturing) model-based process to be used in the reuse and sharing of heterogeneous CAM models created by product manufacturing activities.
4. Construction of knowledge graphs for maritime dangerous goods. (China)	ZHANG, Qi et al. (2019)	Creation of a knowledge *graph of maritime dangerous goods* (KGMDG), to visualize knowledge, retrieval and automatic judgment to promote intelligent transport contributing to the sharing, dissemination and utilization of knowledge.
5. GrEDeL: A Knowledge Graph Embedding Based Method for Drug Discovery from Biomedical Literatures. (China)	SANG, Shengtian et al. (2019)	Creation of the biomedical KG called GrEDeL, which discovers potential drugs for diseases by mining the published biomedical literature.
6. HI2Rec: Exploring Knowledge in Heterogeneous Information for Movie Recommendation. (China)	HE, Ming; WANG, Bo; DU, Xiangkun. (2019)	Creation of KG H12Rec by integrating multiple sources of information to model a recommender system in a unified vector space.
7. A Knowledge Driven Dialogue Model with Reinforcement Learning. (China)	JIA, Yongnan et al. (2020)	Embedding appropriate knowledge into the response generation process, incorporating an end-to-end dialogue model as part of decision-making knowledge as part of response generation.
8. Knowledge graph fusion for smart systems: A Survey. (Coréia do Sul)	NGUYEN, Hoang Long; VU, Dang Thinh; JUNG, Jason. (2020)	Broad, systematic insight into the definitions and challenges of KG fusion in a holistic approach to integrate, enhance, and unify knowledge graphs.
9. Leveraging entity-type properties in the relational context for knowledge graph embedding. (Japão)	RAHMAN, Md Mostafizur; TAKASU, Atsuhiro. (2020)	Knowledge graphs are useful for various artificial intelligence (AI) applications, and the authors have created a KG TPRC *embedding* model for better effectiveness in existing models.
10. Drug repurposing for parkinson's disease by integrating knowledge graph	ZHANG, Xiaolin; CHE, Chao. (2021)	Creation of a KG of medical literature on Parkinson's disease based on local medical knowledge, integrating knowledge into the data in order to obtain better efficacy

completion model and knowledge fusion of medical literature. (China))		in the repurposing of drugs against the disease. Thus, the graph correlates the literature, the drugs applied and the results obtained, favoring the generation of new knowledge and drugs.
11. Leveraging structured biological knowledge for counterfactual inference: A case study of viral pathogenesis. (USA)	ZUCKER, Jeremy et al. (2021)	Study on a causal biological KG analyzing whether the modeling presents learning from the data to better present results through two case studies in the biological area focusing on the type 2 coronavirus.
12. Medical intelligent processor system and traditional Chinese medicine to treat endometriosis. (China)	HAN, Jing et al. (2021)	Creation of the Traditional Chinese Medicine (TCM) KG based on current data from different types of medicines and practices and which is based on the use of technological innovation for knowledge management.

The selected articles show that KG is a multidisciplinary technological solution, used in several sectors, such as health, retail, technology, education, entertainment, among others, for diversified purposes, such as representing and organizing knowledge, pointing out relationships, sharing knowledge, contributing to greater effectiveness of services, etc.

5 CONCLUSIONS

Knowledge management needs to be an integral process, promoting the interaction between information technology, techniques, and people. The use of technological solutions that enable digital environments for collaborative work represents a valuable strategy for companies, enhancing KM.

The present study sought to deepen the understanding of the benefits of using the *knowledge graph* technological solution. With this exploratory study using the Scopus database, it was possible to investigate how academic production is evolving in this term. Even with a low production of academic knowledge, this is still an emerging and important topic for the KM area, due to its interdisciplinary and innovative character. The use of KG in the fields of health, including during the Covid-19 pandemic, in education and in business in general, reinforcing the role of the tool for data processing and knowledge structuring is highlighted.

It is proposed to continue the research, in a systematic and extensive way using other research bases and also to observe the papers presented in national congresses - and in order to verify the application of KG in Brazil - and international, which may bring results that more strongly associate the use of knowledge graphs with knowledge management in the digital age.

REFERENCES

BROWN, S. et al. A academia analítica: uma ponte entre inteligência humana e artificial. 2020. Disponível em < https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-analytics-academy-bridging-the-gap-between-human-and-artificial-intelligence/pt-BR#> Acesso em 01/07/2021.

GARTNER. D&A leaders should understand the key pillars of data fabric architecture to realize a machine-enabled data integration. Mai/2021. Disponível em <https://www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration/> Acesso em 01/07/2021

HE, Ming; WANG, Bo; DU, Xiangkun. HI2Rec: Exploring knowledge in heterogeneous information for movie recommendation. IEEE Access, v. 7, p. 30276-30284, 2019

JIA, Yongnan et al. A Knowledge Driven Dialogue Model With Reinforcement Learning. IEEE Access, v. 8, p. 131741-131749, 2020.

HAN, Jing et al. Medical intelligent processor system and traditional Chinese medicine to treat endometriosis. Microprocessors and Microsystems, v. 82, p. 103842, 2021.

IVANOVA, Rositsa V. et al. How much Knowledge is in Knowledge Graphs?-A Knowledge Management Perspective. 2021 - http://www.semantic-web-journal.net/content/how-much-knowledge-knowledge-graphs-knowledge-management-perspective.

KM World - https://www.kmworld.com/Articles/Editorial/Features/KNOWLEDGE-MANAGEMENT-Industry-Trends-2021-146554.aspx

LI, Xiuling et al. Structured modeling of heterogeneous CAM model based on process knowledge graph. The International Journal of Advanced Manufacturing Technology, v. 96, n. 9-12, p. 4173-4193, 2018.

MOSCHITTI, Alessandro et al. Question answering and knowledge graphs. In: Exploiting Linked Data and Knowledge Graphs in Large Organisations. Springer, Cham, 2017. p. 181-212.

NGUYEN, Hoang Long; VU, Dang Thinh; JUNG, Jason J. Knowledge graph fusion for smart systems: A survey. Information Fusion, v. 61, p. 56-70, 2020.

PAN, Jeff Z. et al. (Ed.). Exploiting linked data and knowledge graphs in large organisations. Heidelberg: Springer, 2017.

RAHMAN, Md Mostafizur; TAKASU, Atsuhiro. Leveraging entity-type properties in the relational context for knowledge graph embedding. IEICE TRANSACTIONS on Information and Systems, v. 103, n. 5, p. 958-968, 2020.

SANG, Shengtian et al. GrEDeL: A knowledge graph embedding based method for drug discovery from biomedical literatures. IEEE Access, v. 7, p. 8404-8415, 2018.

SUCHANEK, F., WEIKUM, G. (2013). Knowledge harvesting in the big-data era. Proceedings of the 2013 International Conference on Management of Data - SIGMOD '13. doi:10.1145/2463676.2463724.

VAN ECK, Nees Jan; WALTMAN, Ludo. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, v. 84, n. 2, p. 523-538, 2010.

VIŠTICA, Marija; GRUBIŠIC, Ani; ŽITKO, Branko. Applying graph sampling methods on student model initialization in intelligent tutoring systems. The International Journal of Information and Learning Technology, 2016.

ZHANG, Qi et al. Construction of knowledge graphs for maritime dangerous goods. Sustainability, v. 11, n. 10, p. 2849, 2019.

ZHANG, Xiaolin; CHE, Chao. Drug Repurposing for Parkinson's Disease by Integrating Knowledge Graph Completion Model and Knowledge Fusion of Medical Literature. Future Internet, v. 13, n. 1, p. 14, 2021.

ZHU, Ji-Zhao et al. Modeling the correlations of relations for knowledge graph embedding. Journal of Computer Science and Technology, v. 33, n. 2, p. 323-334, 2018.

ZUCKER, Jeremy et al. Leveraging structured biological knowledge for counterfactual inference: A case study of viral pathogenesis. IEEE Transactions on Big Data, v. 7, n. 1, p. 25-37, 2021.