Systematic literature review

Technical Report

Draft version 1.0 – March 28th, 2023

Web3 Foundation Grants Program

Contributors

Rafael Brandão (rafael@mobr.ai)
Marcio Moreno (mmoreno@mobr.ai)

1. Introduction

The goal of this systematic literature review is to identify, evaluate, and synthesize available evidence related to the general application of semantic web technologies in Web3, and assess if there is any literature available about this subject referring to the Polkadot ecosystem.

The present systematic literature review follows a predefined protocol, aiming at a thorough and transparent methodology to ensure reproducibility and minimize bias. Figure 1 illustrates the overview of the applied process. It comprises ten steps including, formalizing our research questions, selecting target databases, defining search terms, practical searching on literature, screening the results, qualitative checking regarding selected studies, extracting relevant information, synthesizing information, and finally drawing conclusions based on available evidence.

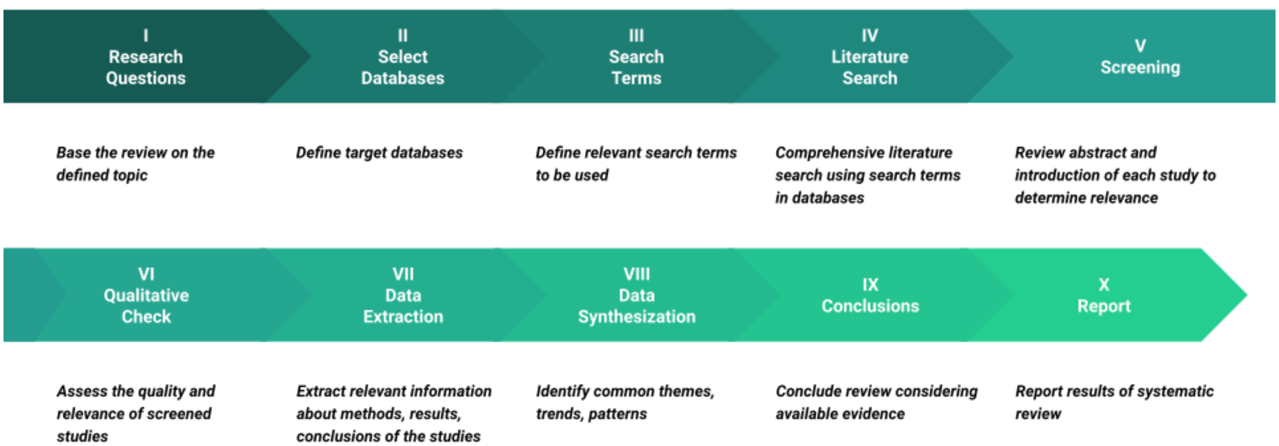

Fig. 1 – Systematic literature review as a 10-step process.

2. Research questions

This literature review is based around the impact of knowledge-oriented approaches (i.e. semantic web technologies) on the development of Web3, and more specifically on the Polkadot multi-chain ecosystem. In this sense, we propose two research questions, one concerning the general aspect of knowledge-oriented approaches and communicability/integration on blockchains, and the other targeting ontologies and Polkadot.

a. General question

How a knowledge-oriented approach may benefit the communicability, integration, usage and development of blockchains?

b. Specific question

How ontology-supported development impacts design and building of Polkadot's parachains and cross-chain communication?

It is important to notice that the research questions proposed here demand long-term research efforts to be properly addressed. The current grant proposal is a first step towards this direction.

3. Databases

To carry the literature search, the research focused on free, open-access databases and metadata search services. We selected four different sources, namely Google Scholar, arXiv, CORE, and DBLP.

a. Google Scholar

Google Scholar is a freely accessible metadata search service that indexes scholarly literature, including articles, theses, books, conference papers, and other documents (both peer-reviewed and not reviewed). It was created by Google in 2004 and provides a simple way to search for scholarly literature across many disciplines and sources, including academic publishers, universities, and preprint repositories. The search results may include citations, abstracts, and full-text articles or documents, depending on availability. The service has an unofficial estimate of over 390M records indexed, as of February 2023.

b. arXiv

arXiv is a repository of electronic preprints of scientific papers in the fields of mathematics, physics, computer science, quantitative biology, quantitative finance, and statistics. It was founded in 1991 and is maintained and operated by Cornell University. The preprints in arXiv are uploaded by authors, and they are not peer-reviewed, though they are moderated for inappropriate content. Access to arXiv is free, and the preprints can be searched and downloaded from their website. As of February 2023, an estimate of over 2.2M records are available in the repository.

c. CORE

CORE (Connecting Repositories) is a search engine and aggregator of open access research papers and theses from repositories and journals worldwide. It aims to increase the visibility, accessibility, and impact of research by providing a single point of access to scholarly outputs. CORE searches through research papers and theses, including pre-prints, from over thousands of repositories and journals, making it a valuable resource for academics. The search service is provided by Knowledge Media Institute and The Open University (UK). As of February 2023, an estimate of 6.5M records are available in the repository.

d. DBLP

DBLP (Digital Bibliography & Library Project) is a digital library and computer science bibliography that indexes and provides open access to major journals, conference proceedings, and other publications in the field of computer science. Its original name was Database Systems and Logic Programming, but it has lost its meaning and was renamed. It was started in 1993 by Michael Ley, and has grown to be one of the most comprehensive databases of computer science literature, with over 6.5M bibliographic records as of February 2023. It is currently maintained by the Leibniz Center for Informatics and University of Trier (Germany). The database is widely used by researchers, students, and practitioners in computer science and related fields. For this reason, we selected the database as a representative of the academia context. Although it sometimes does not provide free access to published materials.

4. Search terms

The literature search considered the combination of keywords listed in Table 1, to cover both the general and specific research questions. It combines relevant aspects of Web3 and knowledge-oriented contexts for the literature search.

Table 1 – Search terms combining Web3 and knowledge-oriented aspects.

Web3 context	Knowledge-oriented context
"web3"	"ontology"
"web3"	"semantic web"
"distributed ledger"	"ontology"
"distributed ledger"	"semantic web"
"blockchain"	"ontology"
"blockchain"	"semantic web"
"polkadot"	"ontology"
"polkadot"	"semantic web"

5. Literature review

As mentioned in Section 3, we prioritized open-access services in our literature search: Google Scholar, CORE, arXiv, and DBLP. The last one (DBLP) is a balancing factor for a more academic-oriented database.

Our methodology consisted on performing queries in the selected databases relating the relevant keywords (Table 1), as well as evident variations about terms, e.g. "web3" or "web 3.0". To summarize the numbers obtained over the search results, we created a series of charts (Figs. 2-5). In addition, details about the search process are documented in a separate spreadsheet[1], with results, further details, and specific query syntax used in searches.

The results show the "blockchain" and "ontology" combination is the most published subject in our search. With the exception of arXiv, in the other three databases it outnumbers by far the other term combinations.

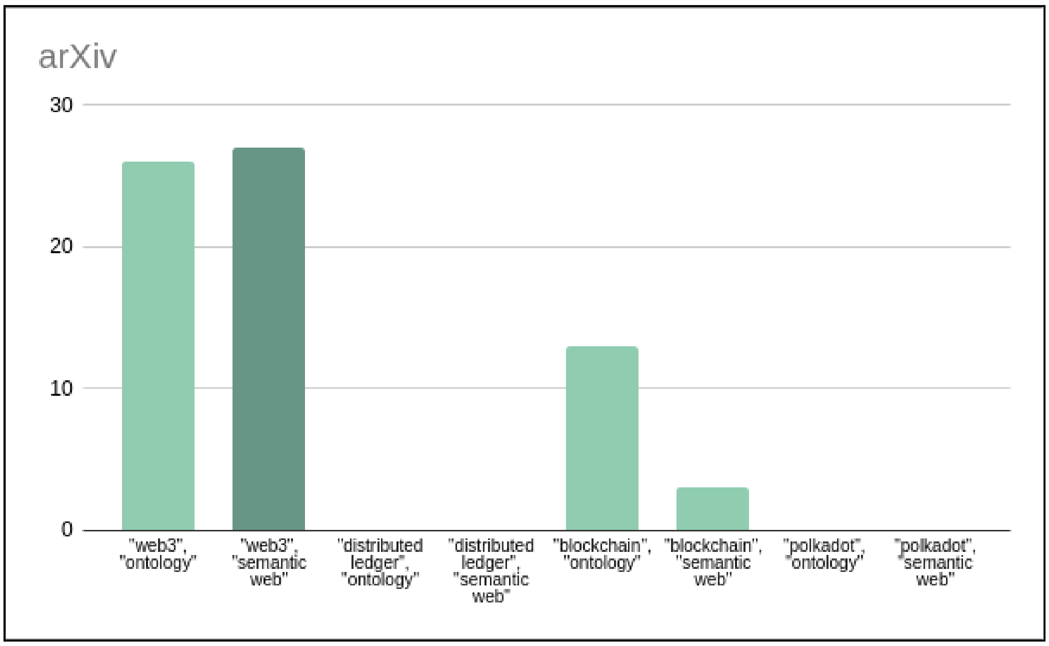

Fig. 2 – Keyword search results in arXiv.

[1] https://github.com/mobr-ai/POnto

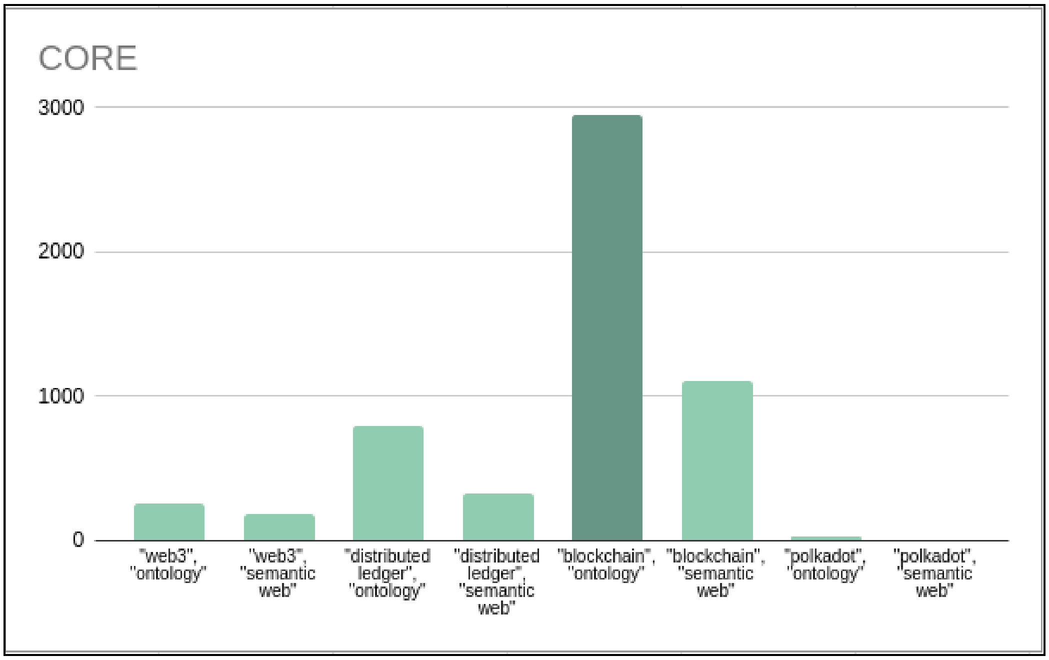

Fig. 3 – Keyword search results in CORE.

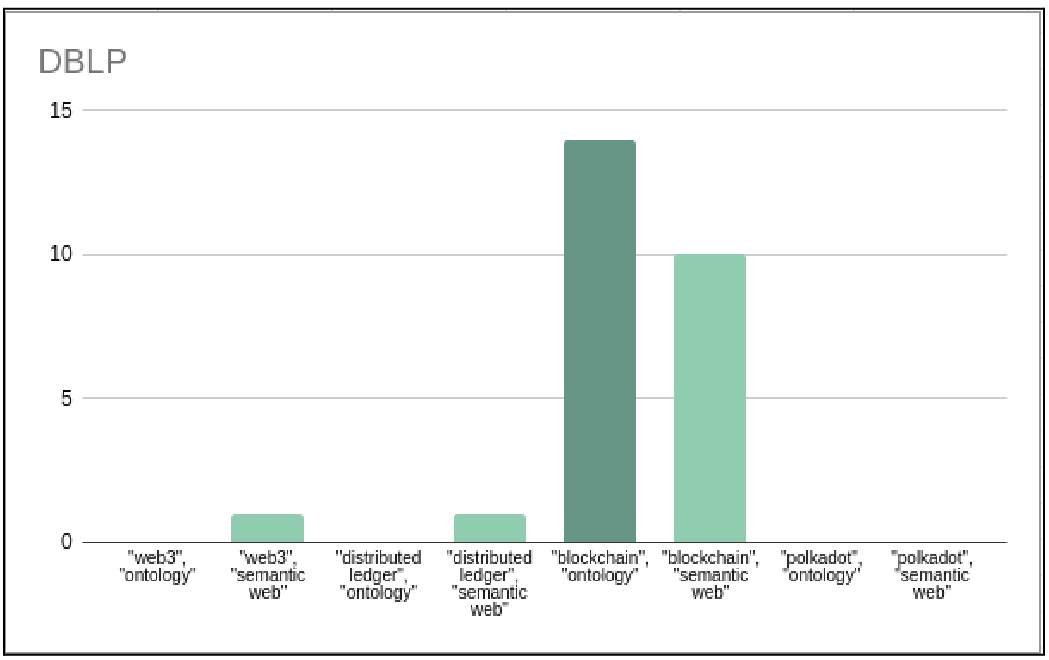

Fig. 4 – Keyword search results in DBLP.

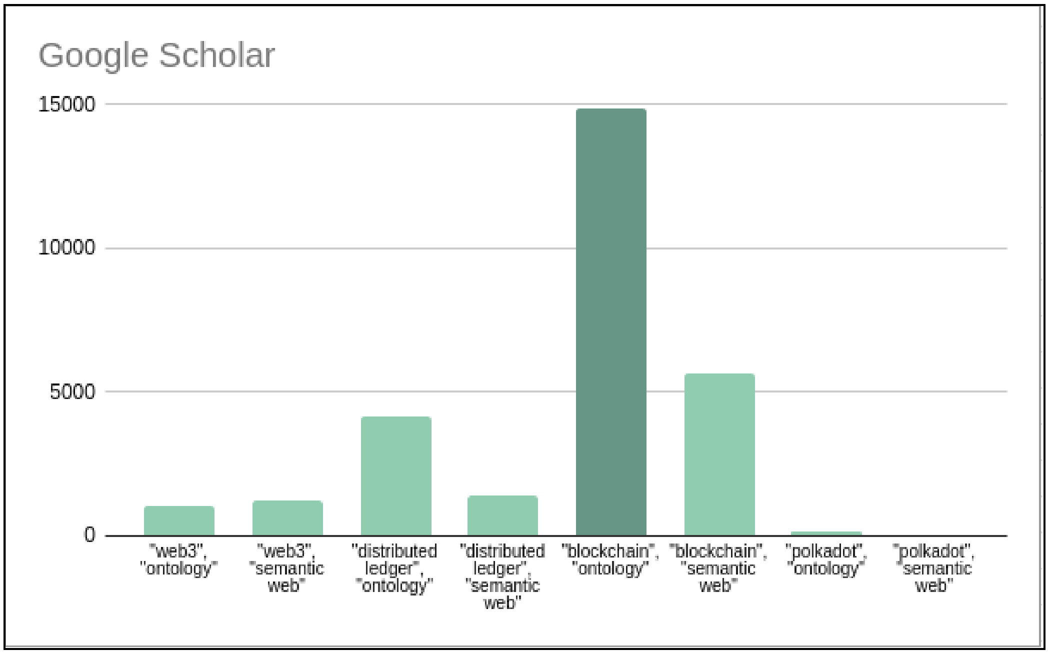

Fig. 5 – Keyword search results in Google Search.

To screen the results, we made the decision to select articles based on their relevance in search ranking. That is, the best ranked results for each keyword combination (first article in query result), from the selected databases. In addition, we considered the top-cited articles and publications for each keyword combination in Google Search. Google Search was selected based on the fact it is the largest database, as well as it is compatible with Harzing's Publish or Perish[2], the tool we used to extract citation metadata. Figure 6 shows a screenshot of the tool during the conducted analysis.

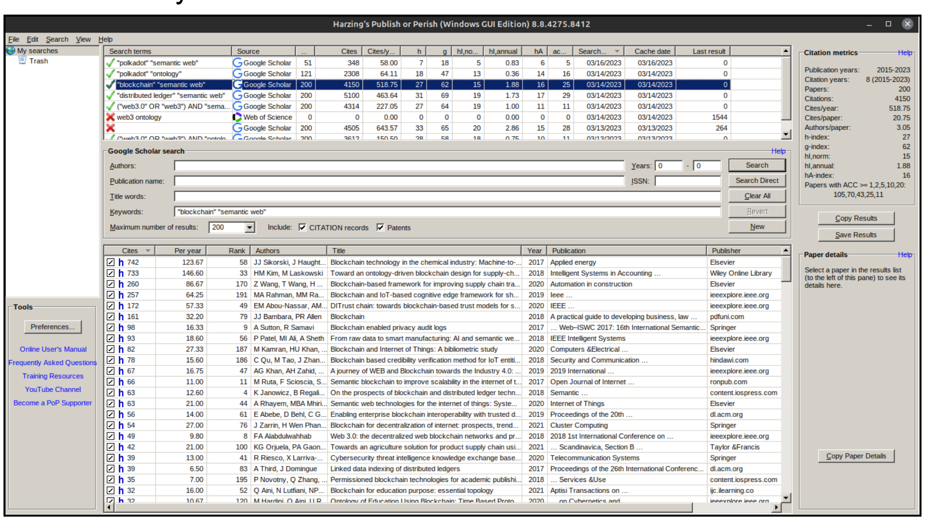

Fig. 6 – Harzing's Publish or Perish tool UI.

[2] https://harzing.com/resources/publish-or-perish/

Tables 2-5 below summarize the best ranked articles per keyword combination. The screening process is documented on a separate spreadsheet structuring the information collected, for further details and information about the process refer to the produced spreadsheet[3].

Table 2 – Results obtained in the arXiv database.

Keywords	Results	Best ranked
"web3", "ontology"	26	Goldston et al., 2023
"web3", "semantic web"	27	Goldston et al., 2023
"distributed ledger", "ontology"	0	NA
"distributed ledger", "semantic web"	0	NA
"blockchain", "ontology"	13	Scrocca et al., 2022
"blockchain", "semantic web"	3	Sheridan et al., 2022
"polkadot", "ontology"	0	NA
"polkadot", "semantic web"	0	NA

Table 3 – Results obtained in the CORE database.

Keywords	Results	Best ranked
"web3", "ontology"	256	Villazón-Terrazas et al., 2009
"web3", "semantic web"	184	Palma et al., 2006
"distributed ledger", "ontology"	790	Laurier et al., 2020
"distributed ledger", "semantic web"	333	Third & Domingue, 2017
"blockchain", "ontology"	2953	Sfetcu, 2019

[3] https://github.com/mobr-ai/POnto

Keywords	Results	Best ranked
"blockchain", "semantic web"	1106	Sfetcu, 2019
"polkadot", "ontology"	32	Besançon et al., 2022
"polkadot", "semantic web"	9	Besançon et al., 2022

Table 4 – Results obtained in the DBLP database.

Keywords	Results	Best ranked
"web3", "ontology"	0	NA
"web3", "semantic web"	1	Bevacqua et al., 2013
"distributed ledger", "ontology"	0	NA
"distributed ledger", "semantic web"	1	Janowicz et al., 2018
"blockchain", "ontology"	14	Besançon et al., 2022
"blockchain", "semantic web"	10	Lin et al, 2023
"polkadot", "ontology"	0	NA
"polkadot", "semantic web"	0	NA

Table 5 – Results obtained in the Google Scholar database.

Keywords	Results	Most cited	Best ranked
"blockchain", "ontology"	14900	Kim et al., 2018	De Kruijff & Weigand, 2017
"blockchain", "semantic web"	5630	Sikorski et al., 2017	Cano-Benito, et al., 2019
"distributed ledger", "ontology"	4130	Kim et al., 2018	Velasco, 2017

"distributed ledger", "semantic web"	1390	Kuo et al., 2017	Janowicz et al., 2018
"web3", "semantic web"	1240	O'Reilly & Battelle, 2009	Jacksi & Abass, 2019
"web3", "ontology"	1020	Kim et al., 2018	Ding et al., 2022
"polkadot", "ontology"	120	Chang et al., 2020*	Besançon et al., 2022
"polkadot", "semantic web"	51	Yang et al., 2019	Abebe et al., 2019

After applying the defined screening strategy, we ended up with 23 articles, including 17 among the best ranked, plus the top-6 most cited from Google Scholar. Table 6 details the six top-cited articles extracted from Google Search using the Publish or Perish tool. Table 7 shows the 17 best ranked articles according to query results in the four databases.

Table 6 – Top-cited articles from Google Scholar.

Title	Author	Type	Venue/Publisher	Citations (Google Scholar)
Blockchain distributed ledger technologies for biomedical and health care applications	Kuo et al., 2017	Journal article	Journal of the American Medical Informatics Association	972
Web Squared: Web 2.0 Five Years On	O'Reilly & Battelle, 2009	Book chapter / Report	O'Reilly Media	929
Blockchain technology in the chemical industry: Machine-to-machine electricity market	Sikorski et al., 2017	Journal article	Applied Energy	742
Toward an ontology-driven blockchain design for supply-chain provenance	Kim et al., 2018	Journal article	Intelligent Systems in Accounting, Finance and Management	733
Blockchain in global supply chains and cross border trade: a critical synthesis of the state-of-the-art, challenges and opportunities	Chang et al., 2020	Journal article	International Journal of Production Research	375
A survey on blockchain-based internet service architecture: requirements, challenges, trends, and future	Yang et al., 2019	Journal article	IEEE Access	96

Table 7 – Best ranked articles among all database searches.

Title	Author	Database	Type	Venue / Publisher	Citations (Google Scholar)
Understanding the blockchain using enterprise ontology	De Kruijff & Weigand, 2017	Google Scholar	Conference proceedings	Int. Conference on Advanced Information Systems Engineering	131
Towards blockchain and semantic web	Cano-Benito, et al., 2019	Google Scholar	Conference proceedings	Business Information Systems Workshops	63
Oyster: sharing and re-using ontologies in a peer-to-peer community	Palma et al., 2006	CORE	Conference proceedings	International Conference on World Wide Web	62
Computing ledgers and the political ontology of the blockchain	Velasco, 2017	Google Scholar	Journal article	Metaphilosophy	48
On the prospects of blockchain and distributed ledger technologies for open science and academic publishing	Janowicz et al., 2018	Google Scholar	Journal article	Semantic Web	40
LinkChains: Exploring the space of decentralised trustworthy Linked Data	Third & Domingue, 2017	CORE	Conference proceedings	Semantic Web	23
Development history of the world wide web	Jacksi & Abass, 2019	Google Scholar	Journal article	Int. Journal of Scientific & Technology Research	18
DeSci Based on Web3 and DAO: A Comprehensive Overview and Reference Model	Ding et al., 2022	Google Scholar	Journal article	IEEE Transactions on Computational Social Systems	10
Web3 Challenges and Opportunities for the Market	Sheridan et al., 2022	arXiv	e-Print	Unpublished / CC BY 4.0	7
A unified blockchain-semantic framework for wireless edge intelligence enabled web 3.0	Lin et al, 2023	DBLP	Journal article	IEEE Wireless Communications	3
A Blockchain Ontology for DApps Development	Besançon et al., 2022	Google Scholar, DBLP, CORE	Journal article	IEEE Access	2
Pattern for re-engineering a classification scheme, which follows the adjacency list data model, to a taxonomy	Villazón-Terrazas et al., 2009	CORE	Conference proceedings	CEUR Workshop Proceedings	2
Traditional accounting with decentralised ledger technology	Laurier et al., 2020	CORE	Conference proceedings	CEUR Workshop Proceedings	1
A Semantic-based Framework for Supporting Interaction and Cooperation in Content-Based Web3.0 Applications	Bevacqua et al., 2013	DBLP	Conference proceedings	Symposium on Advanced Database Systems	0
Blockchain Enterprise Ontologies: TOVE and DEMO	Sfetcu, 2019	CORE	e-Print	Unpublished / CC BY-ND 4.0	0
Digital Inheritance in Web3: A Case Study of Soulbound Tokens and the Social Recovery Pallet within the Polkadot and Kusama Ecosystems	Goldston et al., 2023	arXiv	e-Print	Unpublished / CC BY 4.0	0

Modelling Business Agreements in the Multimodal Transportation Domain Through Ontological Smart Contracts	Scrocca et al., 2022	arXiv	Conference proceedings / e-Print	Int. Conference on Semantic Systems / CC BY 4.0	0

Concerning the most cited articles from Google Search, four of the top-6 focus on blockchain applications in different domains. Kuo et al. (2017) focus on benefits of blockchain for biomedical/health care applications. The fact it is the most cited shows how blockchain is positively seen outside the finance and computing domains. Sikorski et al. (2017) explores application of blockchain in Industry 4.0. Another example of high interest on blockchains outside finance and computing. Chang et al. (2020) explores challenges in global supply chain and trade operations, relevant capabilities and potential applications of blockchain. Yang et al. (2019) explores blockchain-based mechanisms to improve critical features of traditional and centralized Internet services.

Only one of the top cited articles aims at blockchain and semantic web integration. Kim et al. (2018) focuses on developing an ontology-based blockchain for supply-chain provenance. The article provides proof of concept with axioms expressed in first-order logic in smart contracts to execute a provenance trace.

Considering the best ranked publications, 8 out of the 17 articles deal specifically with blockchain and semantic web. De Kruijff & Weigand (2017) propose using Enterprise Ontology (EO) to provide a clear conceptualization of blockchain technology, a four-layered model of blockchain is proposed and applied to a supply chain management system use case. Cano-Benito et al. (2019) proposes a framework for integrating blockchain and semantic web technologies, consisting of three layers: data, semantic, and blockchain. Provides use cases in supply chain management, healthcare, and digital identity management. Third & Domingue (2017) propose a new approach called LinkChains, which combines linked data and blockchain technology. A case study of a LinkChains-based system called Veracity is used for managing data in the agricultural supply chain. Lin et al. (2023) gives an overview of the state of the art in blockchain and semantic web and their applications in "edge intelligence". The publication proposes a unified framework to integrate these technologies and provides a layered architecture with a consensus layer, a semantic layer, and an application layer. The work discusses applications in healthcare, transportation, and smart cities.

Besançon et al. (2022) proposes a blockchain ontology (extending EthOn) with six main components (extending owl:Thing) further divided into sub-components. Authors argue that the ontology can facilitate the development of more interoperable and reusable DApps, reduce development time and costs, and improve the overall quality of applications. They show inference of constraints between the Ethereum blockchain and its sidechain Polygon. Bevacqua et al. (2013) proposes a semantic-based framework supporting cooperation and interaction, including a set of ontology-based services for content annotation, indexing, and retrieval. Sfetcu (2019) presents two ontologies, OMG's Theory of Enterprise Ontology (TOVE) and Design and Engineering Methodology for Organizations (DEMO), as potential solutions for modeling blockchain-based enterprise systems. Scrocca et al. (2022) combines semantic web technologies and smart contracts to enable automated and efficient contract management, execution, and monitoring in the transportation industry. Presents a case study that demonstrates the effectiveness of the proposed framework in managing multimodal transportation business agreements.

To observe trends and patterns, we categorized and explored how the screened publications could be grouped and classified into categories and application domains. Figure 7 shows how they are divided in terms of categories, with most of them (10 publications) being about the application of blockchains in different domains, and another major part (9 publications) considering the integration of blockchains with semantic web technologies. We considered two of the articles to be out-of-scope (O'Reilly Media, 2009 and Jacksi & Abass, 2019). The other three were unavailable for download (Velasco, 2017, Ding et al., 2022 and Bevacqua et al., 2013). Details about this classification can be found in the spreadsheet[4] created during the review process.

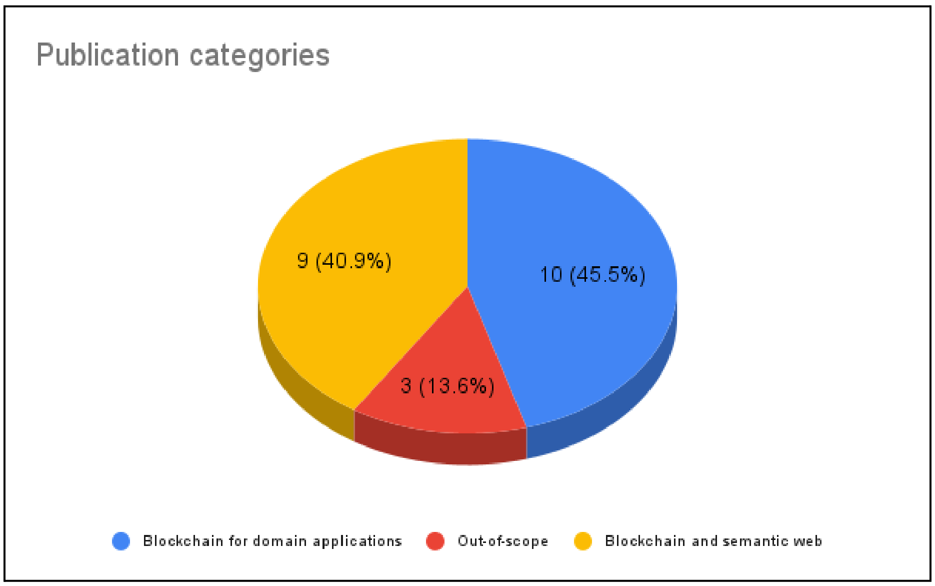

Fig. 7 – Article classification into categories.

Considering the applied domain (Fig. 8), a solid number of articles (5 out of 23) is related to supply-chain. Three of the screened publications are related to ontology development and management. In addition to previously mentioned Besançon et al. (2022) and Sfectu (2019), Palma et al. (2006) presents a decentralized platform for ontology development and sharing. It uses a p2p network, reputation system, and editing tools to facilitate collaboration and promote ontology reuse. This work proposes a decentralized solution, despite preceding the term Web3, coined in 2014 by Gavin Wood.

[4] https://github.com/mobr-ai/POnto

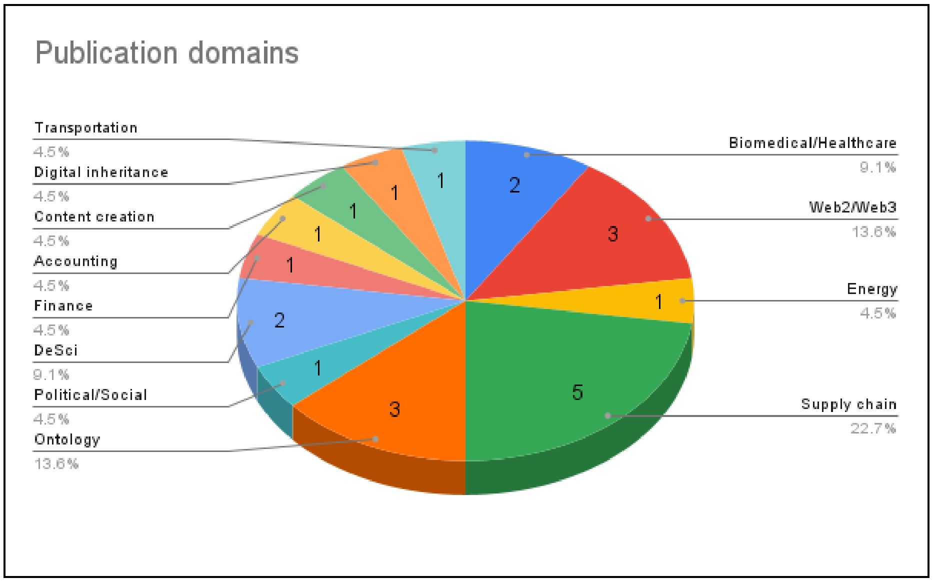

Fig. 8 – Article classification into domains.

6. Final remarks

The top-cited and best ranked perspectives showed how the blockchain and DLTs in general are positively seen by different industry sectors, with a specially tight relationship with supply-chains at large. No surprise here, given the nature of blockchains and the many implementations focusing traceability and trustworthiness aspects.

Another revealing aspect of the conducted literature review is the lack of publications that specifically relate Polkadot with ontologies and semantic web technologies in general. In the conducted search, only three of the screened articles refer to Polkadot. Chang et al. (2020) briefly cites the ecosystem as a viable option tackling the interoperability challenge. Besançon et al. (2022) describes the Polkadot ecosystem as a representative of the multi-chain approach, when discussing principles of communication between shards and multi-chains. Goldston et al. (2023) refers to "ontology" and the Polkadot ecosystem. However, the article cites a project called Ontology, as an option for digital identification. The work describes the Soulbound token and Social Recovery Pallet (SRP) within the Polkadot and Kusama ecosystems as potential solutions for digital inheritance. Meant to overcome the risk of losing access to digital assets and tokens if users pass away or become incapacitated. This is a publication funded by the Web3 Foundation grants program.

References

Abebe, E., Behl, D., Govindarajan, C., Hu, Y., Karunamoorthy, D., Novotny, P., ... & Vecchiola, C. (2019, December). Enabling enterprise blockchain interoperability with trusted data transfer (industry track). In Proceedings of the 20th international middleware conference industrial track (pp. 29-35).

Besançon, L., Da Silva, C. F., Ghodous, P., & Gelas, J. P. (2022). A Blockchain Ontology for DApps Development. IEEE Access, 10, 49905-49933.

Cano-Benito, J., Cimmino, A., & García-Castro, R. (2019). Towards blockchain and semantic web. In Business Information Systems Workshops: BIS 2019 International Workshops, Seville, Spain, June 26–28, 2019, Revised Papers 22 (pp. 220-231). Springer International Publishing.

Chang, Y., Iakovou, E., & Shi, W. (2020). Blockchain in global supply chains and cross border trade: a critical synthesis of the state-of-the-art, challenges and opportunities. International Journal of Production Research, 58(7), 2082-2099.

De Kruijff, J., & Weigand, H. (2017). Understanding the blockchain using enterprise ontology. In Advanced Information Systems Engineering: 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017, Proceedings 29 (pp. 29-43). Springer International Publishing.

Jacksi, K., & Abass, S. M. (2019). Development history of the world wide web. Int. J. Sci. Technol. Res, 8(9), 75-79.

Janowicz, K., Regalia, B., Hitzler, P., Mai, G., Delbecque, S., Fröhlich, M., ... & Lazarus, T. (2018). On the prospects of blockchain and distributed ledger technologies for open science and academic publishing. Semantic web, 9(5), 545-555.

Kim, H. M., & Laskowski, M. (2018). Toward an ontology‐driven blockchain design for supply‐chain provenance. Intelligent Systems in Accounting, Finance and Management, 25(1), 18-27.

Kuo, T. T., Kim, H. E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and health care applications. Journal of the American Medical Informatics Association, 24(6), 1211-1220.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

O'Reilly, T., & Battelle, J. (2009). Web squared: Web 2.0 five years on. " O'Reilly Media, Inc.".

Sikorski, J. J., Haughton, J., & Kraft, M. (2017). Blockchain technology in the chemical industry: Machine-to-machine electricity market. Applied energy, 195, 234-246.

Velasco, P. R. (2017). Computing ledgers and the political ontology of the blockchain. Metaphilosophy, 48(5), 712-726.

W. Ding et al., "DeSci Based on Web3 and DAO: A Comprehensive Overview and Reference Model," in IEEE Transactions on Computational Social Systems, vol. 9, no. 5, pp. 1563-1573, Oct. 2022, doi: 10.1109/TCSS.2022.3204745.

Yang, W., Aghasian, E., Garg, S., Herbert, D., Disiuta, L., & Kang, B. (2019). A survey on blockchain-based internet service architecture: requirements, challenges, trends, and future. IEEE access, 7, 75845-75872.