OpenKG Chain: A Blockchain Infrastructure for Open Knowledge Graphs Postprint

Authors: Huajun Chen, Ning Hu, Guilin, Qi, Haofen Wang, Zhen, Bi, Jie, Li, Fan, Yang, Huajun, Chen

Date: 2022-11-25T00:00:00+00:00

Abstract

The early concept of knowledge graph originates from the idea of the semantic Web, which aims at using structured graphs to model the knowledge of the world and record the relationships that exist between things. Currently publishing knowledge bases as open data on the Web has gained significant attention. In China, Chinese Information Processing Society of China (CIPS) launched the OpenKG in 2015 to foster the development of Chinese Open Knowledge Graphs. Unlike existing open knowledge-based programs, OpenKG chain is envisioned as a blockchain-based open knowledge infrastructure. This article introduces the first attempt at the implementation of sharing knowledge graphs on OpenKG chain, a blockchain-based trust network. We have completed the test of the underlying blockchain platform, and the on-chain test of OpenKG' s data set and tool set sharing as well as fine-grained knowledge crowdsourcing at the triple level. We have also proposed novel definitions: K-Point and OpenKG Token, which can be considered to be a measurement of knowledge value and user value. 1,033 knowledge contributors have been involved in two months of testing on the blockchain, and the cumulative number of on-chain recordings triggered by real knowledge consumers has reached 550,000 with an average daily peak value of more than 10,000. For the first time, we have tested and realized on-chain sharing of knowledge at entity/triple granularity level. At present, all operations on the data sets and tool sets at OpenKG.CN, as well as the triplets at OpenBase, are recorded on the chain, and corresponding value will also be generated and assigned in a trusted mode. Via this effort, OpenKG chain looks forward to providing a more credible and traceable knowledge-sharing platform for the knowledge graph community.

Full Text

Preamble

OpenKG Chain: A Blockchain Infrastructure for Open Knowledge Graphs

Huajun Chen[1]†, Ning Hu[2], Guilin Qi[3], Haofen Wang[4], Zhen Bi[1], Jie Li[2] & Fan Yang[1]

[1]College of Computer Science & AZFT Knowledge Engine Lab, Zhejiang University, Hangzhou 310058, China

[2]Ontology Foundation, Venture Drive #11-31, Vision Exchange, Singapore 608526, Singapore

[3]School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

[4]Intelligent Big Data Visualization Lab, Tongji University, Shanghai 200092, China

Keywords: Open knowledge graph; Blockchain; Decentralized distributed network; Decentralized knowledge graph

Citation: Chen, H.J., et al.: OpenKG Chain: A blockchain infrastructure for Open Knowledge Graphs. Data Intelligence 3(2), 205-227 (2021). doi: 10.1162/dint_a_{00095}

Received: January 10, 2021; **Revised:** March 15, 2021; **Accepted:** April 25, 2021

Corresponding author: Huajun Chen (Email: huajunsir@zju.edu.cn; ORCID: 0000-0001-5496-7442)

Abstract

The early concept of the knowledge graph originates from the semantic Web vision, which aims to use structured graphs to model world knowledge and record relationships between entities. Currently, publishing knowledge bases as open data on the Web has gained significant attention. In China, the Chinese Information Processing Society of China (CIPS) launched OpenKG in 2015 to foster the development of Chinese open knowledge graphs. Unlike existing open knowledge base programs, OpenKG Chain is envisioned as a blockchain-based open knowledge infrastructure. This article introduces the first attempt at implementing knowledge graph sharing on OpenKG Chain, a blockchain-based trust network. We have completed testing of the underlying blockchain platform, as well as on-chain testing of OpenKG' s dataset and toolset sharing and fine-grained knowledge crowdsourcing at the triple level. We have also proposed novel definitions: K-Point and OpenKG Token, which serve as measurements of knowledge value and user value. During two months of testing on the blockchain, 1,033 knowledge contributors participated, with the cumulative

number of on-chain recordings triggered by real knowledge consumers reaching 550,000 and an average daily peak value exceeding 10,000. For the first time, we have tested and realized on-chain sharing of knowledge at entity/triple granularity. Currently, all operations on datasets and toolsets at OpenKG.CN, as well as triples at OpenBase, are recorded on-chain, with corresponding value generated and assigned in a trusted manner. Through this effort, OpenKG Chain aims to provide a more credible and traceable knowledge-sharing platform for the knowledge graph community.

1.1 Knowledge Graphs as World Models

Knowledge—comprising facts, information, or descriptions—represents human awareness and understanding of the world, acquired through experience or education via perception, discovery, and learning [?, ?]. The early concept of the knowledge graph stems from Tim Berners-Lee' s semantic Web vision [?, ?], which aims to model world knowledge using structured graphs and record relationships between entities [?]. Generally speaking, knowledge graphs (KGs) are directed labeled graph (DLG) structures that capture knowledge in the form of triples (subject, predicate, object), expressed as (s,p,o), where s and o denote entities and p establishes their relationship. Due to the convenience of building semantic connections between real-world objects and domain knowledge, many large-scale commercial KGs have been constructed in recent years, such as Google and Baidu Knowledge Graphs, Microsoft Satori, and product knowledge graphs from Alibaba and Amazon. These KGs have enabled a broad range of important applications, including question answering [?], language understanding [?, ?], relational data analysis [?, ?], and recommendation systems [?]. Meanwhile, alongside early AI research on knowledge representations such as ontologies [?, ?] and the consolidation of semantic Web standards like RDF/OWL [?, ?], recent advances in deep representation learning and graph neural networks [?, ?] have spurred new developments in knowledge graph technologies.

1.2 Open Knowledge Graphs

Alongside the burgeoning of the World Wide Web as the largest open knowledge-sharing medium, publishing knowledge bases as open data has gained significant attention since the Web' s early days. Typical examples include the Linked Open Data initiative, which was launched by the semantic Web community and has already collected over 3,000 public linked datasets; ConceptNet [?], which originated from MIT Media Lab' s Web-based, crowdsourced Open Mind Common Sense project launched in 1999; and Wikidata [?], a free and editable knowledge base established by the Wikipedia Foundation. In China, the Chinese Information Processing Society of China (CIPS) launched OpenKG in 2015 to foster the development and openness of Chinese knowledge graphs. OpenKG has accumulated over 200 billion Chinese triples since its inception, with its size growing rapidly. Unlike existing open knowledge base programs, OpenKG Chain is envisioned as a blockchain-based open knowledge infrastructure, which we introduce in detail in the following sections.

1.3 Value Chain of Open Knowledge

Knowledge is a valuable resource. As illustrated in Figure 1 [Figure 1: see original paper], the production, transformation, exchange, and consumption of knowledge form the value chain of knowledge in society. Building a trusted value chain to support the entire knowledge lifecycle upon the open Web infrastructure presents a significant challenge.

1.3.1 Incentive Knowledge at Triple Granularity

The backend logic between contribution and incentives in society can be described vividly: the more contributions we make, the more incentives we receive, motivating people to contribute higher-quality knowledge. According to Maslow' s hierarchy of needs, incentives are not limited to income but should be measurable. As people share their knowledge and create social value, it becomes possible to evaluate and incentivize contributions directly through knowledge itself. For a knowledge graph, another challenge is implementing knowledge-based incentives at triple granularity. From the delivery of factual knowledge in triple format, every step of verification, consumption, transmission, and deletion becomes traceable and measurable. A robust sharing platform must evaluate the value of knowledge in triple form, track contributions during knowledge processing or consumption, and grant proper incentives throughout the knowledge lifecycle.

1.3.2 Self-sovereign Knowledge

Currently, most knowledge graphs use centralized systems where data is stored on centralized servers, preventing contributors from fully controlling their knowledge and transferring ownership to centralized systems. OpenKG Chain uses the term "self-sovereign knowledge" to describe the concept of individuals or organizations taking full responsibility for and control over their knowledge while revealing it with privacy protection and copyright considerations. Individuals and organizations with decentralized identifiers (in any KG system) can discover and locate each other and share knowledge triples without intermediaries, even across systems. OpenKG Chain aims to provide a self-sovereign, interlinking KG system with privacy protection and without intermediaries.

1.3.3 Adversarial Attack and Knowledge Accountability

In an open medium like the Web, everyone can equally publish statements or contribute to and consume knowledge, raising salient concerns about holding all stakeholders accountable for their statements and actions on knowledge. First, statement accountability ensures authenticity when everyone is aware they are responsible for what they state. Second, action accountability monitors all activities performed on knowledge and tracks the entire lifecycle of a triple from its creation through edits, consumption, and deletion, preventing illegal actions or adversarial attacks on the knowledge base. For example, someone might add a malicious rumor to a knowledge base or illegally delete a fact they are reluctant to reveal publicly.

1.3.4 Immutable Knowledge and Tamper Resistance

Another issue relevant to knowledge accountability is tamper resistance. In some cases, knowledge triples are either sensitive or have critical content integrity, requiring protection from fraudulent and intentional modification and necessitating a tamper-resistant network infrastructure. To ensure that once a statement is committed to the network, both its content and all follow-on transactions cannot be altered or compromised retrospectively, the integrity of knowledge content must be guaranteed and transactions must be tamper-proof by any means.

1.3.5 Lighting-up and Dissemination of Knowledge Value

Knowledge consumption provides the most direct way to measure knowledge value—the more knowledge is consumed, the higher its value. Meanwhile, consumption triggers the dissemination of knowledge value, a process we call "lighting-up" of knowledge value. Knowledge graph usage scenarios enable different users to trigger knowledge spread across graph nodes. "Lighting-up by search" occurs when users consume knowledge during the search process, triggering value lighting-up of the searched knowledge item. Knowledge graphs support semantically related search, where further related searches continue to trigger new knowledge lighting-up, with each step recording the generated value. Since knowledge originates from different producers, the produced value must be awarded to corresponding knowledge producers on-chain in an accountable manner. "Lighting-up by question answering" operates similarly: a user' s question triggers lighting-up of the knowledge triple touched by the question, while intermediate nodes traversed from the starting node to the answer node are also lighted up and value-recorded. "Lighting-up by inference" refers to knowledge lighting-up triggered by the inference process. Since knowledge graphs are typically incomplete, reasoning processes complete the graph based on existing knowledge, and due to multiple knowledge sources, lighting-up by inference may also be completed federally—i.e., lighting-up by federal inference [?, ?, ?]. "Lighting-up by analysis" refers to comprehensive analysis of knowledge from different sources continuously triggering lighting processes for related knowledge in the graph, which may also occur in a federated manner due to knowledge source diversity—for example, establishing an analysis model through federated learning [?, ?].

2.1 Blockchain and Distributed Ledger

Blockchain [?, ?] employs distributed ledger technology [?, ?, ?], a ledger database that is shared, copied, and synchronized within an open P2P network where each node participates in data storage and processing. Consequently, each node can monitor transaction legitimacy and testify to transaction results. Blockchain constitutes a multi-centralized network with consensus [?, ?, ?] on a complete transaction log and execution results, characterized by immutability, traceability, and rights confirmation. These characteristics establish a solid "trust" foundation and create a reliable "cooperation" mechanism. An open knowledge platform provides services for collective knowledge maintenance, full tracing of knowledge evolution, quantitative measurement of contributor contributions, and data privacy protection. Knowledge graph development must consider knowledge quantification, iterative history tracing, and governance of knowledge points, contributors, and development environments—requirements that OpenKG blockchain is designed to satisfy.

2.2 Open Knowledge and Blockchain

The construction of open-domain knowledge graphs reflects the social attributes of open communities, presenting various challenges. First, we must identify more individual roles and avoid oligopoly in open knowledge, requiring clarification of how different roles contribute to the open knowledge network. Furthermore, open knowledge contributors must manage their data autonomously to prevent unauthorized misuse caused by data concentration. Second, we need support for more decentralized trust management and more controllable qualification of domain experts for open knowledge. Domain expert qualification is essential for high-quality knowledge crowdsourcing, requiring dynamic adjustment of recognition levels and implementation of more quantitative, fine-grained evaluation programs. Third, we need the ability to quantify contributions from massive participants, tracking the value of open knowledge contributed by many and adjusting the knowledge value model based on feedback from massive participants.

Using distributed ledger technology, the generation, development, and deduction of open knowledge are recorded, enabling full tracking of open knowledge value and ownership. For instance, a multi-centralized blockchain network provides trusted infrastructure, tracks open knowledge development processes, and guarantees data authenticity. A decentralized identity system supports multi-dimensional management of distributed data tokens and massive user tokens. Blockchain' s distributed token solution supports calculation of knowledge value points, reflecting the value of open knowledge.

In summary, open knowledge graphs structured on decentralized distributed networks face many issues, including incentives, ownership management, traceability, trust, and privacy. However, existing centralized knowledge graph management platforms do not address these issues, discouraging knowledge sharing and interconnection while failing to guarantee knowledge authenticity and timeliness. Therefore, we propose a blockchain-based open knowledge graph platform whose functional components can be sorted into three levels: knowledge production, knowledge dissemination, and knowledge consumption, as shown in Figures 2 and 3. The knowledge production layer corresponds to traditional technologies such as knowledge modeling, extraction, fusion, and verification. The knowledge dissemination layer must consider fine-grained knowledge rewarding, self-sovereign knowledge management, knowledge accountability, adversarial attack and tamper resistance, and data privacy protection. The knowledge consumption layer includes semantic search and question answering, reasoning, federated learning, and process automation such as robotic process automation (RPA) [**?**, **?**] and other applications built on distributed knowledge sources.

3.1 OpenKG Resource Model

OpenKG Chain consists of several websites sharing different types of knowledge graph resources. The OpenKG main site provides a sharing platform for coarse-grained open resources such as KG datasets and KG toolsets contributed by China' s KG community. CnSchema provides a crowdsourced open schema for Chinese knowledge graphs. OpenBase is a fine-grained, triple-level knowledge graph crowdsourcing platform. OpenKG Chain has completed construction of the underlying blockchain infrastructure and on-chain testing of sharing different knowledge resource types collected by the OpenKG community. The initial OpenKG blockchain network comprises seven nodes deployed and operated across different universities and corporate institutions. These seven independent nodes form a multi-center blockchain infrastructure for the OpenKG community, built upon a consensus mechanism to provide distributed trusted infrastructure, with more core nodes expandable as needed. On this test platform, over 1,000 knowledge contributors are registered, with a two-month average daily on-chain test value of 10,691 and total lights and on-chain deposits exceeding 550,000. This represents the first test to realize knowledge confirmation at entity/triple granularity (Figure 4 [Figure 4: see original paper]).

3.2 OpenKG Value Model

The first issue OpenKG Chain must address is defining proper value models to reflect knowledge value. For KGs, value calculation requires fine-grained control at the triple level. We propose K-Point to measure knowledge value for triple knowledge published in OpenKG. Secondly, since OpenKG Chain gathers knowledge through community crowdsourcing [**?**, **?**, **?**], we also need a value model to measure and honor knowledge contributor contributions (Table 1).

According to Maslow' s hierarchy of needs, incentives to contribute are not limited to income but should be measurable. We propose a knowledge value point to measure value from a knowledge perspective and OpenKG Token to measure contributions and honor contributors. The initial hypothesis of the knowledge value model is described below and demonstrated in Figure 5 [Figure 5: see original paper].

3.2.1 K-Point: Knowledge Value Measurement

OpenKG Chain designed the K-Point contract to reflect knowledge value. Knowledge value assessment is based on a simple model where each knowledge use increases K-Points accordingly. In current settings, a simple chi-square distribution fits the model as illustrated below. As knowledge usage scenarios increase, OpenKG Chain will continue employing learnable algorithms to calibrate and optimize value evaluation models.

$$f(x; v) = \text{Gamma.dist}(x, v/2, 2)/2$$

Without considering interrelationships of knowledge applications, let:

- K is a single knowledge point
- Knowledge value quota: E_{value}
- Knowledge value expectation: E_{use}
- $(v, 2) = (2, 5)$
- Knowledge value expectations: $E_{\text{value}} = E_{\text{use}}$

Let $x \in (0, 10]$, $\text{Count}_{\text{Knowledge use}}(n)$ be the number of knowledge uses on day n, and the value period of the knowledge point be t (days). Then the unit price of knowledge usage on day n is:

$$\text{Knowledge unit price} = \frac{E_{\text{value}}}{E_{\text{use}}} = \frac{E_{\text{value}}}{(0, 10] \cdot K \cdot \text{Count}_{\text{Knowledge use}}(n)} \cdot E_{\text{use}} \cdot n$$

Each time knowledge is used, K-Points are weighted according to the knowledge usage unit price.

3.2.2 OpenKG Token: Honor Point Measurement

OpenKG Chain designed the OpenKG-Token contract to honor knowledge contributors (publishers, reviewers, and modifiers). OpenKG-Token is dynamically calculated and distributed to contributors when knowledge is used—the more knowledge is used, the more points rewarded to its contributor. Initially, value is equally distributed among knowledge contributors.

$$\text{Single honor value} = \text{Knowledge unit price} \times \text{Count}_{\text{contributor}}$$

The total OpenKG Token satisfies:

$$\text{Total honor value} = \text{Knowledge value} \times g(k)$$

In the initial situation, $g(k) = 1$.

3.3 Decentralized Identity Management in OpenKG Chain

The key to OpenKG Chain is trusted infrastructure. OpenKG Chain adopts the VBFT consensus algorithm, which introduces Verifiable Random Function (VRF) to the traditional Byzantine Fault Tolerance (BFT) algorithm, improving anti-attack capability while increasing consensus speed. Ontology Network uses WasmJIT technology as the smart contract execution environment and provides Layer 2 technologies to balance on-chain business performance and blockchain network expansion solutions (Figure 6 [Figure 6: see original paper]). At the business application level, OpenKG Chain proposes the decentralized identity identification protocol (ONT ID) for identity management throughout the OpenKG Chain lifecycle, including K-Points calculation, resource management, and contributor identification. The distributed data exchange framework (DDXF) manages and tracks the entire knowledge construction, dissemination, and consumption process with cross-system interoperability protocols. ONT ID can issue verifiable credentials for identifying entities, verifying credentials, supporting multi-dimensional authentication, and accessing different trusted sources. Distributed identity identification and multi-dimensional verifiable credentials provide a credible account system and risk control model for different knowledge usage scenarios.

3.4 Data Right Management in OpenKG Chain

OpenKG Chain’ s data construction and use involve multiple rights, including knowledge ownership, sorting, processing, viewing, and downloading. A salient challenge is supporting rights management at different knowledge granularities such as datasets, entities, and triples. OpenKG Chain uses a distributed identity and token scheme to provide fine-grained authority management for multiple knowledge resource types. First, OpenKG Chain' s data—including even a single triple—holds an ONT ID, ensuring unique identification across systems. Second, for different knowledge usage scenarios, knowledge owners and contributors can actively create knowledge authority tokens performed completely on-chain, guaranteeing safety and reliability throughout token usage. Third, all OpenKG Chain users hold ONT IDs, enabling identification of the same user across different knowledge usage scenarios and systems while tracing knowledge contributors across systems to ensure full operation traceability. As shown in Figure 7 [Figure 7: see original paper], implementation details include: (1) Data and user entities have ONT IDs; for different scenarios, addition, deletion, modification, and checking operations on knowledge are managed through off-chain tokens. (2) Each off-chain authority token corresponds to an on-chain data token—OpenKG data-token—using the property relationship between data-token and ONT ID on-chain to confirm cross-system token rights. (3) Operational authentication is performed through the binding relationship between on-chain data-token and off-chain system tokens.

3.5 Trust Management in OpenKG Chain

Why do contributors and users trust knowledge published in OpenKG Chain? For trust management, OpenKG Chain provides credibility metrics for published knowledge at three levels (Figure 8 [Figure 8: see original paper]):

Infrastructure Level. The underlying network scale and node distribution of the OpenKG blockchain provides basic endorsement of published knowledge credibility—it becomes more difficult to cheat on networks with more decentralized nodes.

Knowledge Management Level. Since all operations on knowledge in OpenKG Chain are recorded on-chain in a tamper-proof and traceable manner, they provide trust endorsement for data authenticity and consistency.

Knowledge Contributor and User Level. As all contributor and user behaviors are also recorded and traceable on-chain, behavior analysis can serve as credible endorsement. Notably, while blockchain cannot identify malicious data within the system, it can provide permanently valid proof of malicious behavior outside the system, thereby affecting contributor and user behaviors.

4.1 General Implementation

As previously introduced, the initial OpenKG blockchain network comprises seven nodes deployed and operated across different institutes, forming a multicentered trusted network infrastructure where knowledge contributors can share data and normal users can retrieve knowledge more safely and accountably. For performance considerations, only knowledge operations are recorded on-chain. To synchronize on-chain operation records with off-chain knowledge, OpenKG Chain implements a tokenized contract to solve off-chain knowledge data entity identification problems, recording the entire knowledge token usage process on-chain to ensure operation integrity and traceability. Additionally, OpenKG Chain supports independent knowledge data management by contributors and enables multi-party knowledge collaboration under knowledge privacy protection. In summary, OpenKG blockchain implementation enables new functionalities: (1) Knowledge Index and Resource Synchronization—on-chain records provide an index to off-chain knowledge stored distributively and synchronized through blockchain. (2) Safe Knowledge Consumption—all knowledge consumption types including browsing, downloading, and learning are recorded on-chain, ensuring safer usage and exchange. (3) Accountable Knowledge Processing—all knowledge operations including addition, audit, modification, and abolition are recorded on-chain, ensuring more accountable management. (4) Knowledge Traceability—since all knowledge operations are recorded on-chain, we can trace the changelog of even a single triple based on the alliance chain history.

4.2.1 Introduction to OpenKG.CN

OpenKG.CN is the main website providing a unified sharing platform for different types of open resources, currently supporting sharing of open KG data and open tools where users can freely contribute and download various resources. The platform currently supports three blockchain operations: user registration, resource registration, and resource download. As shown in Figure 9 [Figure 9: see original paper], we built a visualization website for OpenKG community users to check their OpenKG tokens, with real-time resource value updates visible in Figure 10 [Figure 10: see original paper].

4.2.2 Resource Registration on Chain

User registration on-chain. When a user registers at OpenKG.CN, the system automatically completes user information registration on the blockchain server and generates an on-chain account—the ONT ID as the user' s surrogate on-chain.

Resource registration on-chain. After registering at OpenKG.CN, users can upload resources to the platform. For each resource, the system automatically generates a resource ID—the ONT ID for data or tools—and registers the resource on the blockchain server. Note that no OpenKG Token is generated at this stage, as value can only be created when knowledge is consumed.

4.2.3 Resource Value Lighting-up

Since OpenKG.CN provides only coarse-grained knowledge sharing at the dataset/tool level, the main resource consumption and value lighting-up type is implemented through downloading. When resources are downloaded and used by other users, the system generates corresponding OpenKG Tokens according to the resource ID and assigns them to resource contributors' accounts.

4.3.1 Introduction to OpenBase

OpenBase is a crowdsourcing platform enabling fine-grained, triple-level knowledge sharing within the OpenKG community, encompassing triple addition, error checking, and knowledge graph reviewing. OpenBase balances knowledge graph construction cost and speed by combining machine construction with human review and modification. Targeting existing knowledge graphs, OpenBase builds a unified crowdsourcing platform for tasks like error checking and review.

4.3.2 Fine-grained Knowledge on Chain

Traditional knowledge graph crowdsourcing platforms cannot fully solve mutual trust problems among users. Inspired by blockchain, we built OpenBase upon a trusted chain network to enable trusted knowledge sharing at fine-grained triple level. Figure 11 [Figure 11: see original paper] illustrates the crowdsourcing triple procedure on-chain. User operations—including adding new triples, reviewing others' contributions, searching or querying knowledge, or downloading entire datasets—generate related OpenKG Tokens, with all data operations recorded on-chain for future inquiry. For user management, OpenBase registration associates users with ONT IDs decentralized on the underlying blockchain, with all user operations associated with corresponding data and recorded on-chain. For reward management, current settings reward only initial data contributors, not data accessors. OpenKG Tokens are generated when data are reviewed or checked, distributed equally among multiple reviewers and the original contributor. The original uploader remains the data owner, but edit operations make the editor and original contributor co-owners. Data access (search, QA, etc.) constitutes lighting-up operations generating honor points for dataset contributors. Dataset downloads also generate OpenKG Tokens divided among contributors. Adding entities and attributes registers new data, making the operator the new data owner.

4.3.3 Lighting-up Fine-grained Knowledge

For OpenBase, several methods trigger knowledge value lighting-up:

Data search and QA. When users search and query data, corresponding knowledge lights up and generates OpenKG tokens.

Data downloading. Dataset downloads also generate OpenKG tokens to reward contributors.

Data review and checking. Data review and checking operations reward all relevant reviewers or contributors with honor points.

---

5. Conclusion and Future Work

Knowledge is a valuable resource, and linking knowledge further increases its value. The production, transformation, exchange, and consumption of knowledge form society' s knowledge value chain. The knowledge graph value network includes both knowledge contributors and users, with the entire construction and consumption process gradually enriching the knowledge network and essentially increasing knowledge value. This process creates new requirements for triple-granularity knowledge rewarding, self-sovereign knowledge, adversarial attack prevention, and knowledge accountability. Building a trustworthy value chain for the entire knowledge lifecycle upon open Web infrastructure is challenging.

OpenKG Chain addresses these challenges using state-of-the-art blockchain technology, hoping to provide a valuable reference for communities building enterprise-level knowledge graph crowdsourcing platforms. Although blockchain provides new solutions for some problems, it cannot solve all challenges. We still face issues such as performance problems caused by fine-grained knowledge identification on-chain, decentralized knowledge graph storage, and trainable incentive models for knowledge crowdsourcing.

The blockchain-based OpenKG Chain framework provides a technical solution for managing the knowledge lifecycle and its value discovery process from the perspective of knowledge itself. Furthermore, the knowledge formation process faithfully reflects user data self-sovereignty. By bridging physical and digital identities and expanding the knowledge concept to common valuable information from Web pages, forming an Internet "twin"social network becomes possible, which will provide effective experimental support. Currently, OpenKG Chain implements on-chain testing of datasets, toolsets, and triple-form knowledge, with lighting-up methods limited to downloading and searching. Future work will explore more diverse resource types including KG schemas, bots, and knowledge graph algorithms, and richer lighting-up modes such as question answering, decentralized reasoning, and federated knowledge learning.

References

[1] Berners-Lee, T., Hendler, J.: Publishing on the semantic Web. Nature 410, 1023–1024 (2001)
[2] Wang, Z., et al.: Knowledge graph embedding by translating on hyperplanes. In: The 28th AAAI Conference on Artificial Intelligence, pp. 1112–1119 (2014)
[3] Berners-Lee, T., Fischetti, M.: Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor. DIANE Publishing Company, Darby (2001)
[4] Berners-Lee, T., et al.: World-Wide Web: The information universe. Internet Research 20(1), 461–471 (2010)
[5] Berners-Lee, T., Hendler, J., Lassila, O.: The semantic Web. Scientific American 284, 34–43 (2001)
[6] Zhang, Y., et al.: Variational reasoning for question answering with knowledge graph. arXiv preprint arXiv:1709.04071 (2017)
[7] Chen, Y.-N., et al.: Matrix factorization with knowledge graph propagation for unsupervised spoken language understanding. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 483–494 (2015)
[8] Zhang, Z., et al.: Ernie: Enhanced language representation with informative entities. arXiv preprint arXiv:1905.07129 (2019)
[9] Schlichtkrull, M., Welling, M.: Modeling relational data with graph convolutional networks. In: European Semantic Web Conference, pp. 593–607 (2018)
[10] Lin, H., et al.: Learning entity and relation embeddings for knowledge resolution. Procedia Computer Science 108, 345–354 (2017)
[11] Wang, X., et al.: Kgat: Knowledge graph attention network for recommendation. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 950–958 (2019)
[12] Guarino, N.: Formal ontology, conceptual analysis and knowledge representation. International Journal of Human Computer Studies 43, 625–640 (1995)
[13] Gruber, T.R.: The role of common ontology in achieving sharable, reusable knowledge bases. In: Proceedings of KR' 1991, pp. 601–602 (1991)
[14] Guo, Y., Pan, Z., Heflin, J.: Lubm: A benchmark for OWL knowledge base systems. Journal of Web Semantics 3, 158–182 (2005)
[15] Cranefield, S.: Networked knowledge representation and exchange using UML and RDF. Journal of Digital Information 1 (8) (2001)
[16] Velicković, P., et al.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
[17] Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, pp. 1024–1034 (2017)
[18] Liu, H., Singh, P.: Conceptnet—A practical commonsense reasoning tool-kit. BT Technology Journal 22, 211–226 (2004)

[19] Vrandečić, D., Krötzsch, M.: Wikidata: A free collaborative knowledgebase. Communications of the ACM 57, 78–85 (2014)
[20] Coppens, S., et al.: Reasoning over SPARQL. In: LDOW, pp. 1–5 (2013)
[21] Aasman, J.: Utilizing federated knowledge in semantic web applications. In: 2008 IEEE International Conference on Semantic Computing, pp. 486–487 (2008)
[22] Bao, J., Caragea, D., Honavar, V.: A tableau-based federated reasoning algorithm for modular ontologies. In: 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings) (WI' 06), pp. 404–410 (2006)
[23] Konečný, J., et al.: Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492 (2016)
[24] Chen, M., et al.: Fede: Embedding knowledge graphs in federated setting. arXiv preprint arXiv:2010.12882 (2020)
[25] Swan, M.: Blockchain: Blueprint for a new economy. O' Reilly Media, Boston (2015)
[26] Crosby, M., et al.: Blockchain technology: Beyond bitcoin. Applied Innovation 2, 71 (2016)
[27] Ølnes, S., Ubacht, J., Janssen, M.: Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. Government Information Quarterly 34(3), 355–364 (2017)
[28] Mills, D.C., et al.: Distributed ledger technology in payments, clearing, and settlement (2016). Finance and Economics Discussion Series 2016-095. Washington: Board of Governors of the Federal Reserve System. Available at: https://doi.org/10.17016/FEDS.2016.095. Accessed 27 April 2021.
[29] Davidson, S., De Filippi, P., Potts, J.: Disrupting governance: The new institutional economics of distributed ledger technology. Available at: http://dx.doi.org/10.2139/ssrn.2811995. Accessed 27 April 2021
[30] Mingxiao, D., et al.: A review on consensus algorithm of blockchain. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2567–2572 (2017)
[31] Bach, L., Mihaljevic, B., Zagar, M.: Comparative analysis of blockchain consensus algorithms. In: The 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1545–1550 (2018)
[32] Nguyen, G.-T., Kim, K: A survey about consensus algorithms used in blockchain. Journal of Information Processing Systems 14(1), 101–128 (2018)
[33] van der Aalst, W.M., Bichler, M., Heinzl, A.: Robotic process automation. Business & Information Systems Engineering 60(4), 269–272 (2018)
[34] Willcocks, L.P., Lacity, M., Craig, A.: The IT function and robotic process automation. LSE Research Online Documents on Economics 64519, London School of Economics and Political Science, LSE Library (2015)
[35] Yang, J., Adamic, L.A., Ackerman, M.S.: Crowdsourcing and knowledge sharing: Strategic user behavior on taskcn. In: Proceedings of the 9th ACM conference on Electronic commerce, pp. 246–255 (2008)
[36] Chu, X., et al.: Katara: A data cleaning system powered by knowledge

bases and crowdsourcing. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, pp. 1247–1261 (2015)
[37] Roy, S.B., et al.: Task assignment optimization in knowledge-intensive crowdsourcing. The VLDB Journal 24, 467–491 (2015)

Author Biography

Huajun Chen is a Full Professor in the College of Computer Science and Technologies at Zhejiang University, serving as Director of the Joint Lab on Knowledge Engine of the Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies (AZFT) and Deputy Director of the Key Lab of Big Data Intelligence in Zhejiang Province. His research interests include knowledge graph technologies, ontologies and the semantic Web, information extraction and natural language processing, and applications in biomedicine, e-commerce, and smart cities. ORCID: 0000-0001-5496-7442

Ning Hu is Chief Technology Officer at Ontology Foundation, a high-performance public blockchain specializing in decentralized identity and data. He has nearly two decades of experience in software development and high-end technology, including semantic Web, big data, and artificial intelligence (AI). ORCID: 0000-0002-9598-8554

Guilin Qi is a Professor at Southeast University, China, where he serves as Director of the Institute of Cognitive Intelligence and of the Knowledge Science and Engineering Lab. His research interests include knowledge representation and reasoning, knowledge graphs, uncertainty reasoning, and the semantic Web. Prof. Qi is an editorial board member of the Journal of Web Semantics and has co-edited special issues for the Annals of Mathematics and Artificial Intelligence, International Journal of Approximate Reasoning, and Journal of Applied Logic. With over 20 years of research experience in knowledge engineering, he has led many national and industrial projects on knowledge graphs, published more than 100 papers, holds two patents, and is one of China' s leading knowledge graph experts. ORCID: 0000-0002-1957-6961

Haofen Wang has long served as Chief Technology Officer in a first-line artificial intelligence company with rich experience in AI R&D management. He is one of the founders of OpenKG, the world' s largest Chinese open knowledge graph alliance, and has participated in numerous provincial and ministerial AI-related projects. He has published over 90 high-level AI papers cited more than 2,000 times (H-Index: 21), built the world' s first interactive virtual idol "Amber · Xuyan," and developed an intelligent customer service robot serving over 1 billion users. He currently serves as Deputy Director of the Terminology Committee of China Computer Society, Deputy Secretary General of the Language and Knowledge Computing Committee of Chinese Information Society, and Distinguished Research Fellow in the College of Design Innovation at Tongji University. ORCID: 0000-0002-0672-8081

Zhen Bi is a graduate student at Zhejiang University. His research interests include natural language processing and knowledge graph. ORCID: 0000-0002-3287-5683

Jie Li is currently a Back-end Developer at Ontology Foundation, devoted to developing data and identity information based on blockchain technology.

Fan Yang is a graduate student at Zhejiang University. His research interests include natural language processing and knowledge graph. ORCID: 0000-0002-4577-8196

Note: Figure translations are in progress. See original paper for figures.