Toward Blockchain-Assisted Gamified Crowdsourcing for Knowledge Refinement

Helun Bu(✉) and Kazuhiro Kuwabara

College of Information Science and Engineering,
Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan
is0385pr@ed.ritsumei.ac.jp

Abstract. This paper reports an application of blockchains for knowledge refinement. Constructing a high-quality knowledge base is crucial for building an intelligent system. One promising approach to this task is to make use of "the wisdom of the crowd," commonly performed through crowdsourcing. To give users proper incentives, gamification could be introduced into crowdsourcing so that users are given rewards according to their contribution. In such a case, it is important to ensure transparency of the rewards system. In this paper, we consider a refinement process of the knowledge base of our word retrieval assistant system. In this knowledge base, each piece of knowledge is represented as a triple. To validate triples acquired from various sources, we introduce yes/no quizzes. Only the triples voted "yes" by a sufficient number of users are incorporated into the main knowledge base. Users are given rewards based on their contribution to this validation process. We describe how a blockchain can be used to ensure transparency of the process, and we present some simulation results of the knowledge refinement process.

Keywords: Blockchain · Knowledge refinement · Gamified crowdsourcing

1 Introduction

This paper describes an application of blockchains for a knowledge refinement process. Constructing a high-quality knowledge base is important for an intelligent system. Knowledge refinement is necessary to increase the value of the knowledge base. Several approaches to refining knowledge represented as knowledge graphs have been reported [13].

One promising approach is to harness the power of many users, for example, through crowdsourcing [6]. The inherent problem in crowdsourcing is maintaining a high quality output. It is important to motivate users (or workers) to produce robust results.

Gamification is one way to keep users motivated in crowdsourcing [12]. Gamification is the process of introducing game-like elements to a non-gaming context. Games with a purpose (GWAP) is an example of gamification, where intended tasks are executed as by-products of playing games [2]. GWAP is also applied to

N. T. Nguyen et al. (Eds.): ACIIDS 2020, LNAI 12033, pp. 3–14, 2020.
https://doi.org/10.1007/978-3-030-41964-6_1

refine knowledge graphs [14]. In the ESP game, which is a prominent example of GWAP, two users who do not communicate with each other are asked to label an image. Points are given to a user who puts the same label as a paired user [1]. By devising proper game rules, an incentive can be given to users to input correct answers.

Introducing a point system is expected to motivate users to earn more points and complete more tasks. In such situations, correctly calculating points is important. In particular, when points cannot be calculated at the time users are playing the game or completing the task and are calculated later according to the user's past contribution, it is necessary to ensure that the users' records that are used for points calculation are not altered in any way. This point calculation process needs to be transparent so that any user can examine its basis. For this purpose, we utilize blockchains, which is a distributed ledger technology that has been proposed as the basis of cryptocurrency [3]. Blockchains allow the data to be stored and shared over the network, with a guarantee of being free from tampering.

In this paper, we focus on the example of the knowledge base used in our word retrieval assistant system [9]. This system is intended to support people with word-finding difficulties. Through a series of questions and answers, the system tries to guess what the person wants to express but cannot find a name for. The knowledge base is used to formulate questions.

For this system, we accumulate knowledge contents from various sources such as scraping websites or obtaining inputs from human users. Because of the nature of this system, its knowledge content should cover topics that often appear in everyday conversation. They are generally not related to specialized domain knowledge; rather, they are about things related to daily life. Thus, the participation of many casual users to construct the knowledge base would be effective.

However, when many casual users contribute to knowledge content, the quality of the knowledge may become an issue. To ensure high quality knowledge, we first store newly acquired pieces of knowledge into a temporary knowledge base, and only move the validated knowledge into the main knowledge base.

In this validation process, we employ a concept of gamified crowdsourcing similar to the one explored in [8]. More specifically, we make simple yes/no quizzes from the contents of the temporary knowledge base, and we present these to users. When enough votes for agreeing with the content are accumulated, the corresponding knowledge content is judged to be valid, and is incorporated into the main knowledge base. As an incentive to users, we award points based on a user's past inputs. If they contribute to the validation process, they are given bonus points as rewards. Blockchain technology is utilized to record the users' input and ensure the transparency of the reward calculation.

The remainder of the paper is organized as follows. The next section describes some related work focusing on applications of blockchains for handling knowledge, and Sect. 3 presents the knowledge refinement process. Section 4 describes the prototype implementation, and Sect. 5 reports simulation experiments and discusses their results. Section 6 concludes.

2 Related Work

Blockchains were developed as an underlying technology for cryptocurrency and were intended for use as the public ledger for transactions on a network. Blockchains are now applied not only to cryptocurrency, but also to other areas such as health care, data provenance, and mobile communication networks [11]. One notable application area is the Internet of Things (IoT), where blockchains are utilized to construct a *knowledge market* in which IoT systems that perform artificial intelligence (AI) tasks at the edge of networks can exchange knowledge in a peer-to-peer fashion [10]. For knowledge management in enterprises, *Knowledge Blockchain* was proposed [5] to audit knowledge evolution and provide proof of provenance of knowledge.

A blockchain is also used as a decentralized database where data from participants are stored transparently. Knowledge graphs represented in the Resource Description Framework (RDF) are stored using a blockchain technology, called GraphChain [15]. In addition, blockchains are applied to the decentralized construction of knowledge graphs [16]. In this system, company-level domain knowledge about employee's skills is constructed from the participation of employees in the company. This system also introduces a voting scheme and a reward mechanism for employees who contribute to the knowledge construction. As another example, AUDABLOK was proposed as a software framework to allow citizens to participate in refining open data [4]. In AUDABLOK, blockchains are utilized to audit users' contributions and provide rewards to users.

In this paper, we focus on the validation process of knowledge contents by casual users and conduct simulation experiments to examine the method's characteristics.

3 Knowledge Refinement Process

3.1 Target Knowledge Base

As an example knowledge base, we chose the knowledge content used in our word retrieval assistant system [9]. A typical problem for people with aphasia is word-finding difficulty; they have a clear image of what they want to say, but cannot recall a proper word to express it. This is similar to the situation when you visit a foreign country and you do not know how to say in local language.

For this kind of difficulty, a human caregiver called a conversation partner often asks a series of questions, such as *Is it food?* or *Is its color red?* Through their responses to the questions, the conversation partner extracts the name of the thing the person with aphasia wants to express.

The word retrieval assistant system aims to provide a similar function to a human conversation partner, but using a computer. The system contains a knowledge base about relevant things and produces an appropriate question to ask the user. According to the user's reply, the next question to ask is determined.

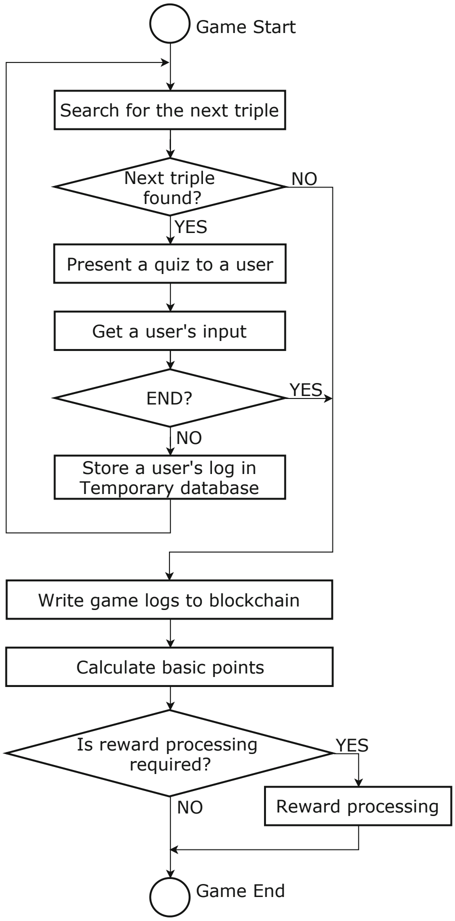


Fig. 1. Flowchart for a game session.

In the early prototype system we developed, knowledge is represented as a triple of *subject*, *predicate*, and *object* as in the RDF. For example, the fact that *the color of the apple is* red is represented as `(<apple>, <color>, <red>)`.

3.2 Knowledge Acquisition and Refinement

It is important to acquire enough knowledge for the word retrieval assistant system to work well. Techniques to acquire new knowledge include system developers constructing knowledge content manually from scratch or letting a user input the correct word when it is not produced by the system [7].

In addition, we may extract knowledge from data available on the internet, such as Wikipedia. Alternatively, we may use a framework of crowdsourcing to elicit inputs from many casual users. To facilitate such a process, gamification

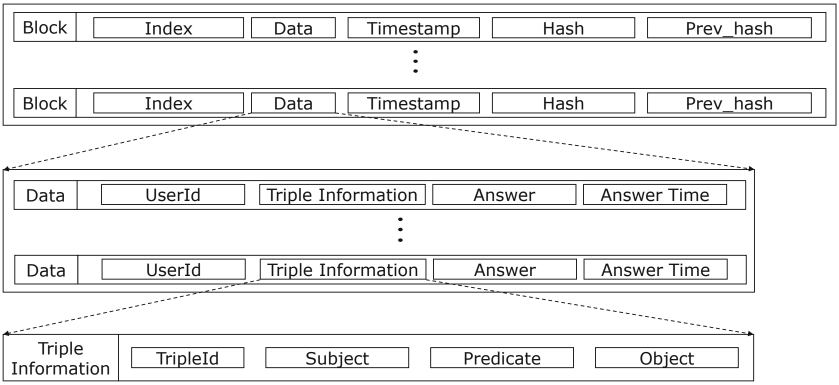


Fig. 2. Data stored in blockchain.

concepts can be applied. For example, when we need data for a triple, we may present a form that consists of three items, one or two of which are blank so that a user can fill them in.

One of the problems with data acquired using such methods is that they may not be correct. To assure high quality of knowledge content, we validate each triple's correctness before they are used. The acquired knowledge content is first stored in the temporary knowledge base. Only the knowledge content validated by users is moved into the main knowledge base.

To validate the contents of the temporary knowledge base, we employ a yes/no quiz. For example, suppose the triple `(<apple>, <color>, <red>)` is in the temporary knowledge base. A yes/no quiz is presented to a user asking if the color of the apple is red is correct or not. The user may answer YES, NO, or DON'T KNOW. If a certain number of YES votes compared with NO votes are obtained, the triple is considered to be true and is moved into the main knowledge base.

This process is formulated into a kind of game. After a user starts a game, a yes/no quiz sentence is presented with the possible choices: YES, NO, DON'T KNOW, and END. If the user answers END, the game session ends. Otherwise, a user's reply is recorded in the game server, and the next quiz is presented with the same possible choices (Fig. 1).

When one game session is finished, the ending processing is performed. This includes storing logs of the game session into blockchain by creating a new block (Fig. 2). The game history consists of triples that correspond to quizzes that were presented to the user and the user's responses to them. It also includes the date and time that the user played the game.

Furthermore, basic points are given to the user for each of their answers. Finally, possible bonus rewards are calculated, as explained in the next subsection.

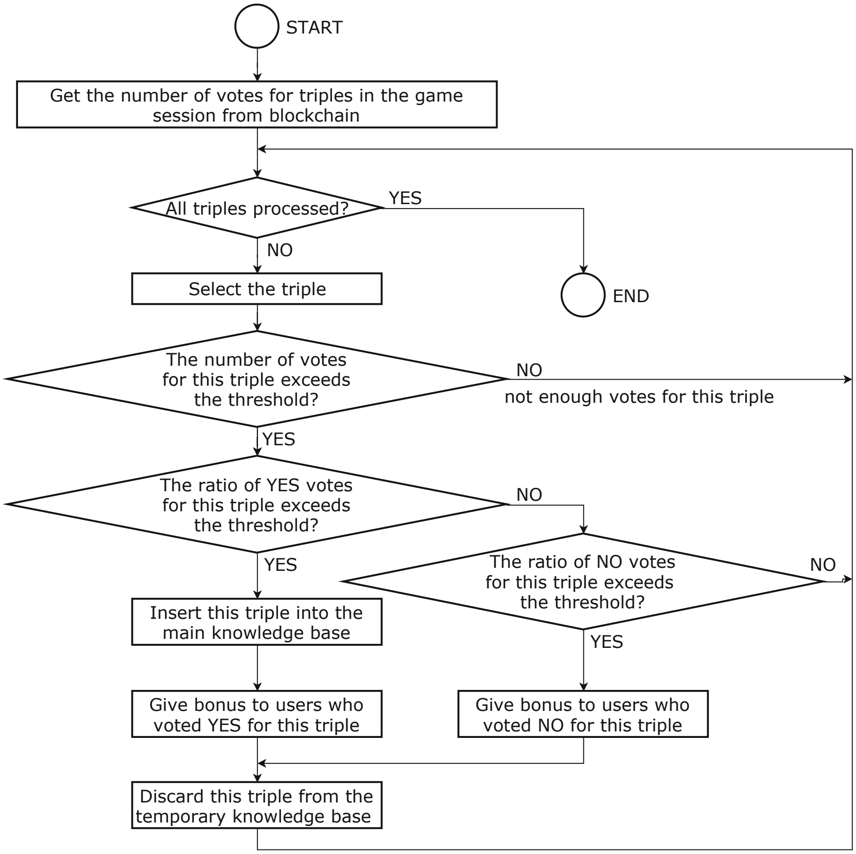


Fig. 3. Reward processing.

3.3 Reward Processing

For triples that are used in the game session, we check whether enough votes (YES or NO) from the game playing are obtained for the triple (Fig. 3). We set a certain threshold S, and check whether a triple has accumulated more than S votes. The data are obtained from the blockchain so that transparency of data is ensured. If the triple has accumulated more than S votes and the ratio of YES votes over the total number of votes for the triple is greater than a threshold α, then the triple is judged to be correct. The triple is removed from the temporary knowledge base and moved into the main knowledge base. Similarly, if the ratio of NO votes over the total number of votes for the triple is greater than a threshold β, the triple is judged to be incorrect and is removed from the temporal knowledge base.

When a triple is judged to be correct, the users who voted YES for this triple are given bonus points. Similarly, when a triple is judged to be incorrect,

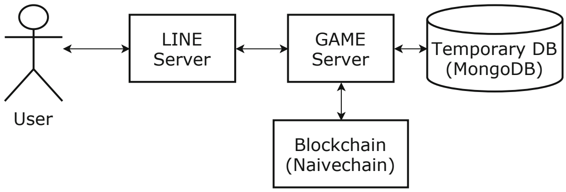
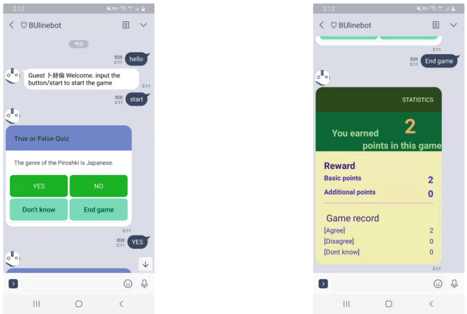


Fig. 4. Prototype configuration.



(1) During a game session (2) After a game session

Fig. 5. Screenshots of a prototype chat system.

the users who voted NO for this triple are given bonus points. This check is done for all the triples that appeared in the game session.

4 Implementation

Figure 4 shows the configuration of our early prototype system, which is constructed as a chat system. We implemented it using LINE messaging service[1], a popular chat service in Japan and other countries. A user can access the system which is implemented as a chatbot, via a smartphone so that it can be used even in small periods of free time. When the user starts talking to the chatbot, the chatbot presents a quiz that displays the sentence generated from a target triple and buttons (YES, NO, DON'T KNOW) to input the user's response along with a button to end a game (Fig. 5(1)). When the game ends, the summary of the game results are shown (Fig. 5(2)).

We used the implementation of Naivechain[2] for blockchain. A MongoDB server is used to store the temporary data generated during a game session, and also for simulation experiments explained in the next section. Each time a game session ends, a list of the user's inputs during the game session is sent to the Naivechain server using an HTTP POST request, and the session data are added

[1] https://line.me/en/.
[2] https://github.com/lhartikk/naivechain.

to the blockchain. Before rewards processing is performed, the games' logs are retrieved from the Naivechain server using an HTTP GET request.

5 Experiments

5.1 Purpose and Methods

To examine the characteristics of the proposed method with some variations of the game design, we conducted simulation experiments rather than probabilistic analysis. In addition, as our early prototype with the particular implementation of a blockchain has performance issues in terms of scalability, the simulations were performed using MongoDB as the data store.

We assumed M triples and N virtual users, and let virtual users participates in the games. We examined whether the triples can be properly validated and how much rewards virtual users receive.

More specifically, the conditions of the simulation are as follows:

- The number of triples (M) was set to 1000, and the number of virtual users (N) was set to 1000.
- User i ($1 \leq i \leq N$) answers correctly with probability $p_c(i)$, where $p_c(i)$ is uniformly distributed over $[0.6, 1.0]$. Thus, the average is 0.8. In addition, a user is assumed not to answer with DON'T KNOW.
- All the triples are assumed to be true. Thus, user i answers with YES with probability $p_c(i)$. Otherwise, user i answers NO.
- A user is selected from a set of N users in sequence, and the selected user plays a game. The user is assumed to answer k quizzes in one game session. The value of k is set randomly from the range of $[3, 7]$. The user is given 1 point per quiz as a basic point.
- As one game session finishes, we check whether there are newly validated triples. The threshold for judging a triple to be correct or not (S) is set to 10, and the threshold ratio for judging the triple to be correct (α) is set to 0.8. That is, a triple that accumulates 10 votes, among which more than 8 votes are for YES, is considered correct. The users who voted YES for the triple are given 100 bonus points. After triples to be validated are determined and bonus points are calculated, the next user is selected and another game is started.
- One simulation run is terminated when there are no more triples to be presented to any user. Note that the same triple will not be presented to the same user twice. Once a triple is removed from the temporary knowledge base, it is not used for a game.

Triple Selection Method. We tested three variations in selecting the next triple to use for a yes/no quiz. When the next triple is searched for, the triples that have already been used for the target users are removed first. Then, from the remaining triples, we used the following methods to select the next triple to present.

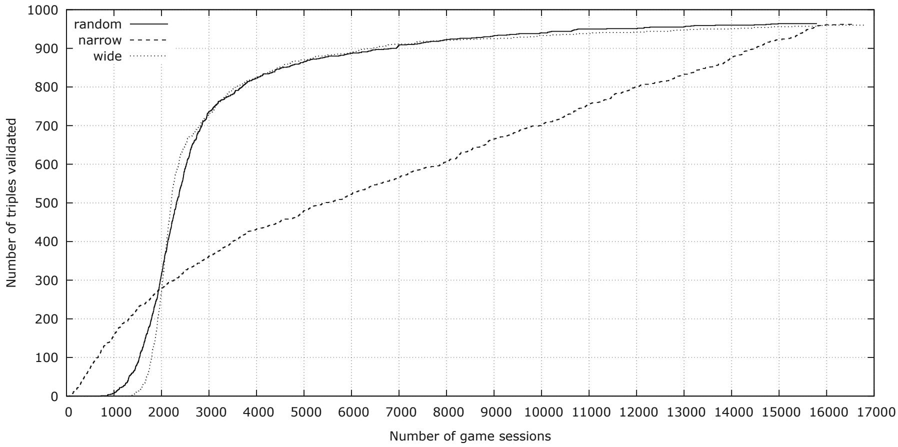


Fig. 6. Changes in the number of triples validated and the number of game sessions in one simulation run.

- *random*: A triple is selected randomly; this is the baseline method.
- *narrow*: A triple that has been presented to users more often is given higher priority. This method effectively focuses on some triples to validate. More specifically, each triple is given a weight equivalent to 10 times the number of times it has been used as a question to users.
- *wide*: A triple that has been presented to users less often is given higher priority. This method effectively broadens the range of target triples. More specifically, each triple is given a weight equivalent to 10 times the maximum number of times any triple has been used as a question minus the number of times the particular triple has been used as a question.

User Reliability. We also ran another set of simulations under different conditions to account for user reliability. Each user's reliability is calculated based on their contribution to the validation of triples. The threshold S for determining whether a triple is valid is set according to the sum of the reliability of users who voted for the triple so that fewer users are required to validate a triple when users with high reliability vote for it. This process is expected to result in fewer games required for validation.

For the simulation condition, the reliability of user i, R_i is set as $R_i = \min(0.002 \times b_i + 1, 3)$ where b_i denotes user i's bonus points. Thus, the range of R_i is $1 \leq R_i \leq 3$, meaning that a user of high reliability counts as, at most, three users.

5.2 Results and Discussion

Simulations were run five times without considering user reliability. Across the simulation runs, the average number of validated triples for each method was

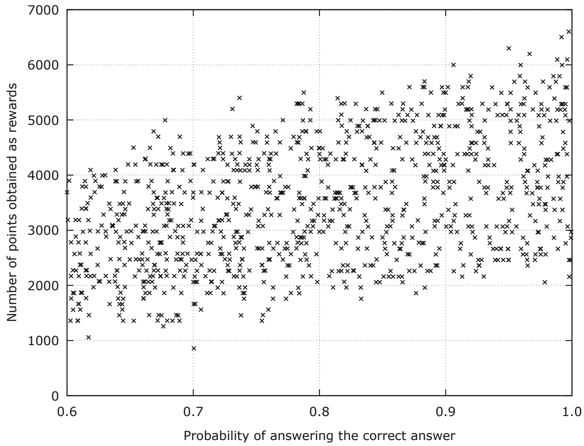


Fig. 7. Relationship between probability of answering a correct answer and points obtained.

967 (*random* method), 966.2 (*narrow* method) and 964 (*wide* method). Generally, this kind of yes/no quizzes can validate enough triples.

The different triple selection methods impacted how the number of triples that are were validated through the games changes, as shown in Fig. 6. As seen in this chart, after a certain number of games, the number of validated triples rapidly increases and then saturates for the *random* and *wide* methods. Among the approaches, the *wide* method is slow to validate triples, but the number of validated triples gradually increases compared with the *random* method.

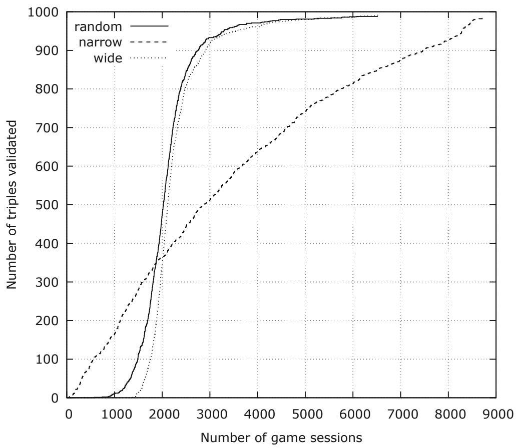


Fig. 8. Changes in the number of triples validated and the number of game sessions (with user reliability considered).

Although the *narrow* method shows a steady increase in the number of validated triples, it is soon surpassed by the other methods.

We also plotted the number of points user i obtained against the user's probability of giving the correct answer ($p_c(i)$) in one simulation run with the *random* method (Fig. 7). The correlation coefficient between was 0.45, indicating a moderate positive relationship. Generally, a user who has a higher probability of answering a correct answer tends to obtain more points as expected.

In addition, the simulation results with user reliability taken into consideration are shown in Fig. 8. As seen in this chart, the number of game sessions required to validate triples were decreased by introducing user reliability. With the parameter settings in the simulation runs, there was no adverse effect such as erroneously validating triples. It is a future task to confirm that introduction of user reliability does not cause any adverse effect in a more realistic situation.

6 Conclusion and Future Work

This paper described the use of blockchain in knowledge refinement. We adopted the concept of gamified crowdsourcing and used blockchains to ensure transparency of the user reward calculation. Our simulation experiments indicate that the proposed approach has the potential to be used as a method of knowledge refinement.

In this paper, we focused on the process of validating the knowledge content already in the knowledge base. We plan to extend the application of blockchains to the process of gathering the knowledge content.

Furthermore, we conducted only simulation experiments with virtual users. Blockchains were introduced to ensure the transparency of the user reward calculation so as not to hinder the users' incentives. Future work should evaluate the proposed method not only in terms of performance scalability in real-world environments, but also in terms of human users' subjective impressions.

Acknowledgements. This work was partially supported by JSPS KAKENHI Grant Number 18K11451.

References

1. von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, pp. 319–326. ACM, New York, April 2004. https://doi.org/10.1145/985692.985733
2. von Ahn, L., Dabbish, L.: Designing games with a purpose. Commun. ACM **51**(8), 58–67 (2008). https://doi.org/10.1145/1378704.1378719
3. Antonopoulos, A.M.: Mastering Bitcoin: Programming the Open Blockchain. O'Reilly Media, Sebastopol (2017)
4. Emaldi, M., Zabaleta, K., López-de Ipiña, D.: AUDABLOK: engaging citizens in open data refinement through blockchain. Proceedings **31**(1), 52 (2019). https://doi.org/10.3390/proceedings2019031052

5. Fill, H.G., Häerer, F.: Knowledge blockchains: applying blockchain technologies to enterprise modeling. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018). https://doi.org/10.24251/HICSS.2018.509
6. Howe, J.: Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. Crown Business, New York (2009)
7. Iwamae, T., Kuwabara, K., Huang, H.H.: Toward gamified knowledge contents refinement - case study of a conversation partner agent. In: Proceedings of the 9th International Conference on Agents and Artificial Intelligence. ICAART 2017, vol. 1, pp. 302–307 (2017)
8. Kurita, D., Roengsamut, B., Kuwabara, K., Huang, H.H.: Simulating gamified crowdsourcing of knowledge base refinement: effects of game rule design. J. Inf. Telecommun. **2**(4), 374–391 (2018). https://doi.org/10.1080/24751839.2017.1401259
9. Kuwabara, K., Iwamae, T., Wada, Y., Huang, H.-H., Takenaka, K.: Toward a conversation partner agent for people with aphasia: assisting word retrieval. In: Czarnowski, I., Caballero, A.M., Howlett, R.J., Jain, L.C. (eds.) Intelligent Decision Technologies 2016. SIST, vol. 56, pp. 203–213. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39630-9_17
10. Lin, X., Li, J., Wu, J., Liang, H., Yang, W.: Making knowledge tradable in edge-AI enabled IoT: a consortium blockchain-based efficient and incentive approach. IEEE Trans. Industr. Inf. **15**(12), 1 (2019). https://doi.org/10.1109/TII.2019.2917307
11. Lu, Y.: The blockchain: state-of-the-art and research challenges. J. Ind. Inf. Integr. **15**, 80–90 (2019). https://doi.org/10.1016/j.jii.2019.04.002
12. Morschheuser, B., Hamari, J., Koivisto, J., Maedche, A.: Gamified crowdsourcing: conceptualization, literature review, and future agenda. Int. J. Hum. Comput. Stud. **106**(Supplement C), 26–43 (2017). https://doi.org/10.1016/j.ijhcs.2017.04.005
13. Paulheim, H.: Knowledge graph refinement: a survey of approaches and evaluation methods. Semant. Web **8**(3), 489–508 (2017)
14. Re Calegari, G., Fiano, A., Celino, I.: A framework to build games with a purpose for linked data refinement. In: Vrandečić, D., et al. (eds.) The Semantic Web - ISWC 2018, pp. 154–169. Springer, Cham (2018)
15. Sopek, M., Gradzki, P., Kosowski, W., Kuziski, D., Trójczak, R., Trypuz, R.: GraphChain: a distributed database with explicit semantics and chained RDF graphs. In: Companion Proceedings of the The Web Conference 2018, WWW 2018, pp. 1171–1178. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2018). https://doi.org/10.1145/3184558.3191554
16. Wang, S., Huang, C., Li, J., Yuan, Y., Wang, F.: Decentralized construction of knowledge graphs for deep recommender systems based on blockchain-powered smart contracts. IEEE Access **7**, 136951–136961 (2019). https://doi.org/10.1109/ACCESS.2019.2942338