Exploration of Blockchain Technology and IPFS Distributed Storage System in UGC Model

Zhexi Gu[1], Lijun Yang[2], Baolin Liu[1]

zhexi_gu.lzu@outlook.com (Zhexi Gu); jyang@njupt.edu.cn (Lijun Yang); sun8000680@outlook.com (Baolin Liu)

[1] School of Journalism and Communication, Lanzhou University, Lanzhou 730000, China
[2] School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract: The model of UGC has been widely used in content production on Internet platforms, breaking the previous model of professional producers and institutions monopolizing content production. However, the UGC model also faces difficulties in verifying the sources of information, the dissemination of false news, copyright protection, and confirmation. By taking advantage of the traceability of blockchain technology, this paper explores a new model of information release, dissemination, verification, and re-dissemination based on blockchain technology, and provides corresponding implementation methods. In terms of copyright protection, the NFT technology, and IPFS application are used to directly confirm the copyrights of user-produced content and generate its unique corresponding digital ownership certificates.

Keywords: UGC, Blockchain, NFT, IPFS, information verification, copyright protection

1 Introduction

1.1 The concept and development of UGC

*U*GC (User Generated Content) was first introduced in 2005 by Mary Meeker, a principal analyst at Morgan Stanley, and the rise of UGC is inextricably linked to the new Internet revolution of Web 2.0. Compared with Web 1.0, the most important feature of Web 2.0 is that users can generally participate in the construction of website content, not only "readable", but also "writable". Where the Web 1.0 consisted of a one-way traffic from producers of content to a mostly passive public, the Web 2.0 involves a broad interaction between traditional content producers on the one hand, and more active users, who are participating, commenting, and also generating content themselves on the other hand [1]. Instead of simply getting and downloading information from the Web, UGC is now produced and uploaded. Users not only become the receivers of web content, but also the producers of web content. Regarding the definition of UGC, the 2007 report of the Organization for Economic Cooperation and Development (OECD) proposed three characteristics of UGC: Internet-based open, innovation, and creation by non-professionals or non-authoritarians [2]. With the development of UGC platforms such as YouTube video sites and TikTok, more and more creators with professional backgrounds and MCN (Multi-Channel Network) organizations have joined in the production of UGC. Therefore, some scholars suggest that UGC can be defined as anything that meets the three characteristics

ICEMBDA 2023, October 27-29, Tianjin, People's Republic of China

DOI 10.4108/eai.27-10-2023.2341906

of being uploaded to the Internet platform by users, being available for viewing and sharing by non-closed collection members, and having or partially having its own opinions and ideas [3].

In the future, by using blockchain, a decentralized technology to distribute and store data under Web 3.0, users will no longer be dominated by the Internet centralized platform, and they will have stronger autonomy. In the meantime, as the scale of Internet users continues to expand, the Internet infrastructure continues to improve, and with new technologies such as 5G empowerment, the threshold for users to enter the field of content production will continue to lower, and the UGC content production model will certainly be further developed.

1.2 UGC platform and content production

The application of UGC platforms with values co-created as the core concept has been quite extensive. Values co-created by users and the platform, e.g. pragmatic value, hedonic value, and user equity, further pro- mote the generation of users' behavioral intention of making continuous contributions [4]. In foreign countries, social platforms such as Wikipedia, YouTube video sites, Twitter and Facebook have been quite mature in their UGC production mode. In China, short video platforms represented by Douyin and Kuaishou, Q&A and knowledge-sharing platforms represented by Douban and Zhihu, and WeChat and Weibo platforms with social attributes have been formed. In addition, some Chinese news platforms are also actively exploring the UGC model, such as the "Wenba" section of Pengpai News.

According to the motivation and nature of the content uploaded by users, the content of UGC platforms can be roughly divided into three types: first, social content mainly for personal records and sharing; second, information with certain news value uploaded by users in the place where the event occurred, such as videos taken by eyewitnesses at the accident scene; and third, commercial promotion placed on the platform by businesses or enterprises in the form of individuals. Among them, the first category of social content and the third category of advertising content have a less public impact on society, while the second category of news content is often easy to become a news hotspot or a source of rumors because it contains an amount of information that the public cares about and can be widespread forwarded and discussed over the Internet.

2 The problems faced by the UGC production model

2.1 Difficulty in information verification

The development of UGC has changed the original mode of publishing information by professional media, and its characteristics of fast update, large quantity, and wide range make UGC not only become the main source for audiences to get information but even become the main source for professional media to find hot news and get first-hand on-site information.

The 50th Statistical Report on the Development Status of the Internet in China released by the China Internet Information Center (CNNIC) shows that as of June 2022, the size of China's Internet users reached 1.051 billion [5]. Few people seem to check the reliability of news before sharing them with their friends and potentially with millions of others. This is mainly due to the fact that the Internet, in particular social networking services, provides a complete decentralization of information on a large scale: every user is potentially a news source, and

often it is not trivial to establish the truth[6]. The large base of Internet users in China produces a wide variety of content, and the manual audit cannot cope with the huge volume of content, which inevitably leads to the weakening of the gatekeeping link. Moreover, the anonymity of the Internet platform and the casual presentation also exacerbate the difficulty of verifying information.

2.2 Fake news

Users under the UGC model, unlike professional media in the past, have relatively low media literacy. Additionally, the concealment of the Internet makes it difficult to verify the authenticity of information sources, and the lack of encryption makes news products at risk of malicious tampering [7]. Furthermore, some organizations and users specialize in fabricating fake news with sensational effects to increase traffic and attention. Currently, social media platforms represented by Weibo, WeChat and Douyin have become the main way for people to exchange information and self-expression, and the more convenient and efficient information dissemination has also contributed to the spread of fake news [8]. Moreover, although each platforms have gradually established mechanisms for dispelling fake news, when the spread of fake news is far greater than that of debunking messages, fake news will continue to be circulated and shared on the platform.

2.3 Difficulties in copyright protection

On October 30, 2018, the Beijing Internet Court held a public hearing on the first case it accepted after its inauguration --A case of copyright ownership and infringement dispute between "Douyin Short Video" and "Huopai Small Video". The short video involved in the case was only 14 seconds long and was independently created by the creator "Blackface V" with personalized expression. The court held that it should be protected as a work under copyright law [9].

In addition to legal issues, UGC infringement also faces problems such as difficulty in protecting rights, low cost of infringement, and slow processing of review, resulting in the protection of copyright in UGC being particularly difficult.

3 The application of blockchain technology in the UGC problem

The concept of blockchain was first proposed by Satoshi Nakamoto in 2008[10]. Blockchain is an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way [11]. All transaction data in the blockchain is organized into an ordered block list, and the submitted block information can no longer be tampered with in any way [12].

A block consists of a version number, timestamps, a random number, the hash of the previous block, Merkle Root, and a target hash. Each block forms a chain structure in order, which is also known as a blockchain ledger. When a new block is generated, the hash and random number of the block need to be recalculated based on the hash value of the previous block, Merkle root, and the random number. The Figure 1 shows Merkle Root.

Blockchain will provide powerful technical support to address the aforementioned difficulties in the UGC model.

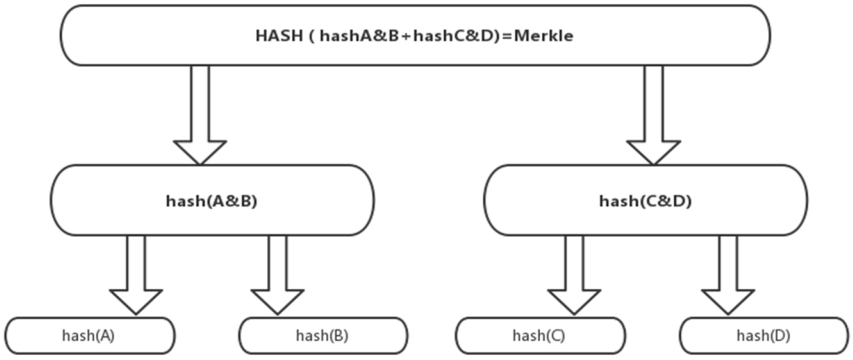

Figure 1 Merkle Root

3.1 "Timestamp" traceability - help information verification and news hot spot discovery

Most of the traditional traceability technologies rely on intermediate carriers, such as adding QR codes, RFID, NFC tags, or direct laser coding to the items on the product packaging. However, due to the problem of the trustworthiness of source information in these practices, the asymmetry of source information makes it possible for any stakeholder to modify the data information on its traceability chain at any time, and its cost is very low. The emergence of blockchain technology undoubtedly provides another optimal solution for transaction data traceability authentication technology. Traceability in blockchain technology is achieved through distributed ledger, consensus mechanisms, and immutability. A timestamp, on the other hand, is a record of the time and order of transactions and other information in the blockchain system. Each transaction in a blockchain has a timestamp, and these transactions are grouped together in a block, which also has a timestamp. These timestamps serve as a proof of when the transaction took place. The transactions and timestamps are encrypted and linked to the previous block in the chain, creating a tamper-proof, chronological record of all transactions on the blockchain. This allows for easy tracking and verification of the history of transactions, making the blockchain an ideal technology for applications requiring traceability and transparency. The figure 2 indicates the concept of Blockchain.

When the content uploaded by users to the platform reaches a certain threshold of the impact of its reprints and comments, professional media will be introduced to directly find the source of information through the timestamp of the blockchain and verify the content.

For the calculation of the heat threshold of UGC content, this paper adopts the following formulas.

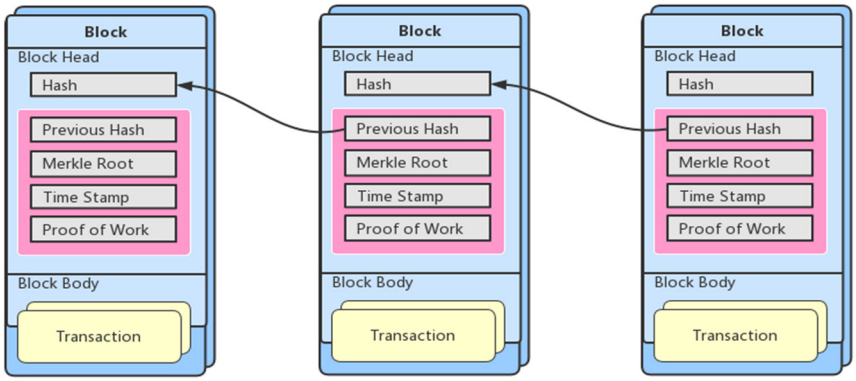

Figure 2 Blockchain

3.1.1Quantifying the initial impact of users

Based on the macro-level Social Network Analysis, the relationship between a single node and other nodes is directed and there are strong and weak relationships, sometimes interacting with each other. The initial model G(V, E) is obtained, where V represents the set of user nodes in the social grid and E represents the relationship type of these nodes.$\boldsymbol{n_1, n_2, n_3, n_4, n_5, n_6}$..... with connected aeolotropism of the grid model, and a random wandering between these directed connected nodes [13]. In this paper, we use the PageRank algorithm combined with the depth and breadth of user influence to quantify the user's initial influence:

$$User\_V_0 = \sum_{n \in N} \frac{UGC\_V_n}{L(n)} \tag{1}$$

Where n is all the individual nodes that interact with the target node, all n are included in N, $L(n)$is the number of outer nodes of node $n.UGC_V_n$is the initial influence assignment score of UGC content in different domains.

3.1.2Initial impact assignment of UGC content in different fields

The quantitative analysis was performed using Naive Bayes model. Assuming A_i is the i-th UGC field, A is the set of domain categories $A=\{A_1, A_2, A_3, A_4, A_5 \ldots\ldots\}$. B_i is the i-th valid classification term of a specific UGC field, $B = \{B_1, B_2, B_3, B_4, B_5 \ldots\ldots\}$, Bayesian formula: $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)}$. Based on the criterion of mutual independence of each variable, the Bayes classification processor estimates the probability of occurrence of UGC content in different fields and the probability of occurrence of keywords in UGC content through learning and then calculates the probability of belonging to different fields for each combination of keywords, and finally outputs with the maximum probability [14].

$$\prod_{i=1}^{n} P(B_i|A) = P(b_1|A) * P(b_2|A)P(b_3|A) \ldots P(b_n|A) \tag{2}$$

$$UGC\_V_0 = k \prod_{i=1}^{n} P(B_i|A) \tag{3}$$

Where $P(B_i|A)$ denotes the UGC content in theB_i the i-th categorized word in B_i the probability of appearing in the set A, k is the scale factor of different fields.

3.1.3Users re-disseminate impact

In this paper, we use the SIR (Susceptible-Infected-Recovered) model to calculate the re-dissemination impact of users, where S stands for users who have not received the information but have the ability to spread it, and I stands for users who have received information and are in the process of dissemination, Rstands for users who cannot receive the information or are unwilling to disseminate it. Assuming the parameter α is the dissemination efficiency,β is the inefficiency, N is the total number of users (assuming no change), the unit of time is t, γ is the influence coefficient. The following relationship can be obtained.

$$N(t)=S(t)+I(t)+R(t) \tag{4}$$

$$I(t)=\alpha S(t) \tag{5}$$

$$R(t)=\beta I(t) \tag{6}$$

S,I,R are all continuous derivative functions with respect to t, the differential equation is found as

$$\frac{dS}{dt} = -\alpha S(t)I(t) \tag{7}$$

$$\frac{dI}{dt} = -\beta I(t) + \alpha S(t)I(t) \tag{8}$$

$$\frac{dR}{dt} = \beta I(t) \tag{9}$$

The influence of re-dissemination is a continuous function of time, integrating over the variables can be organized as follows

$$RS\_V_0 = \gamma \int_{t=i}^{n}((\alpha + 1)S(t) - R(t))d(t) \tag{10}$$

The social impact of a UGC is calculated by summing the individual variables and multiplying by a time decay factor.

$$UGC\_V = (User\_V_0 + UGC\_V_0 + RS\_V_0) * Time\_F \tag{11}$$

Time_F is decayed according to the e^{-x}. The UGC impact function value exceeds a threshold value and shows a rapid increase, but will eventually converge to 0 again with time. Therefore, the final selected UGC impact formula is as follows.

$$\left[\sum_{n\in N}\frac{UGC\_V_n}{L(n)} + k\prod_{i=1}^{n}P(B_i|A) + \gamma\int_{t=i}^{n}((\alpha+1)S(t) - R(t))d(t)\right] * e^{-x} \tag{12}$$

The preset threshold is the maximum value corresponding to the first time the UGC_V function approaches the surge point.

3.2 Super nodes mark - cut off the chain of fake news dissemination

After reaching the threshold, professional media intervene in the process of information dissemination as a super node, verify the content uploaded by users, mark it as true or false, and

give it a true or false stamp, greatly simplifying the previous process of investigation, debunking, dissemination of debunking information. In particular, the marked content will display the true or false stamps both before and after the marking. While reducing the cost of repeated verification by other media, it also makes it impossible for rumors to continue to spread due to the inadequate and untimely spread of debunking messages in the past. Instead, every time a message marked as false is disseminated, it can be seen as a debunking. The figure 3 depicts the process of verifying and re-disseminating information.

After verifying the authenticity of the information, the professional media further explores and expands its news value, thus forming a UGC+PGC production model.

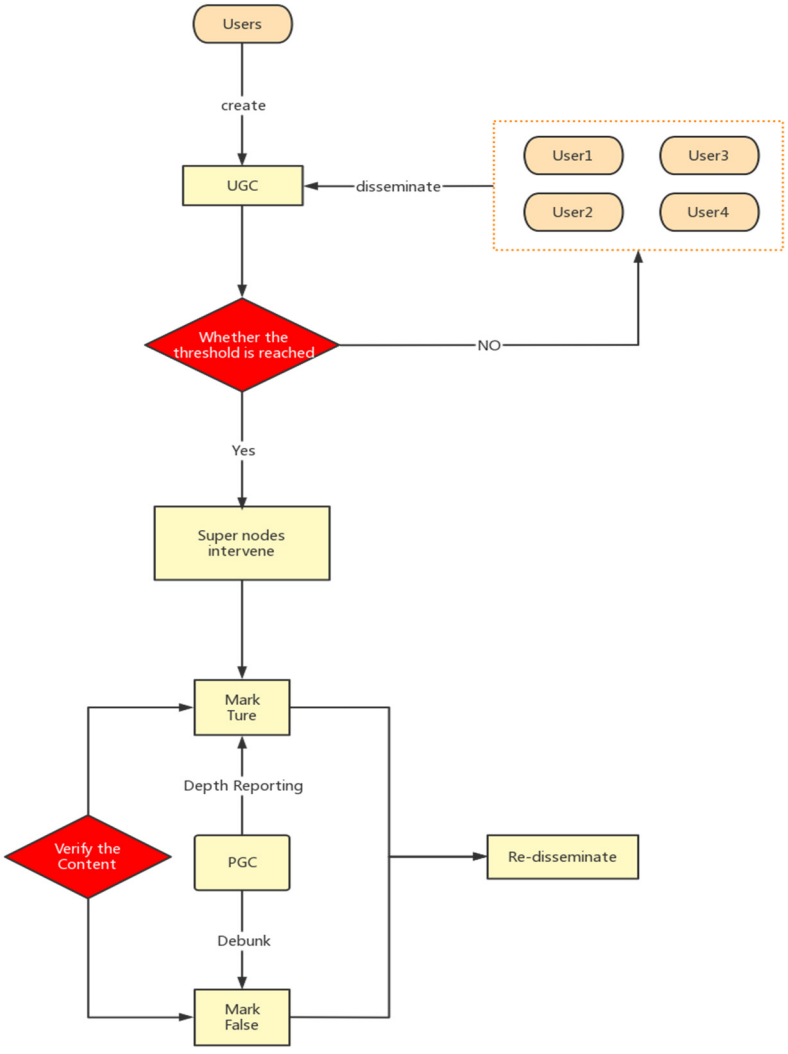

Figure 3 Flowchart of "Dissemination - Verification - Re-dissemination"

3.3 NFT confirmed copyrights - to protect interests of content producers

In March 2021, several news media organizations began testing the NFT market. Quartz sold a story as an NFT, claiming it was the first news story to be sold as an NFT [15]. The New York Times, Time Inc., and Hearst Media Group are among the U.S. media giants that have entered the NFT market. In December 2021, Xinhua News Agency released a limited edition of selected photojournalism from 2021 in the form of China's first digital news collection, marking the official entry of the "national team" of China's news media organizations into the NFT digital collection. This marks the official entry of the "national team" of China's news media organizations into the NFT digital collection market [16].

The key to dealing with copyright issues lies in the confirmation of rights. NFT (full name Non-Fungible Token) is a cryptocurrency with blockchain as the underlying architecture, and thus has features such as data structure, distributed storage, encryption algorithm, and consensus mechanism [16]. Anyone can obtain copies of digital items (NFTs) on the blockchain, but they are tracked to provide proof of ownership [17]. The content uploaded by users becomes digital assets through "NFTization" and is stored on the blockchain, with unique rights credentials and records of transactions and modifications attached to it, to be "confirmed the copyrights". Users can view and download these contents by paying some tokens. The traditional membership system and paid subscription model will also be better used in the UGC production model so that the rights and interests of content producers are fully protected.

3.4 IPFS - Distributed Storage System - Record on-chain assets

IPFS (InterPlanetary File System) is a distributed file storage system whose most important feature is that it is not the old HTTP protocol of IP addressing, but content addressing instead. Each file uploaded to the IPFS system is cut into small fragments. Each small fragment is individually hashed once, and the encrypted values are eventually combined and hashed again. At the same time, the IPFS system stores these fragments in each node, and when these files need to be called, the system retrieves these slices from the nodes closer to the network and restores them for use [18]. Compared to a single blockchain system, the introduction of IPFS can provide more on-chain storage space and significantly improve the efficiency of communication between nodes.

3.5 Decentralized UGC publishing platform

This paper attempts to propose a decentralized UGC publishing platform based on IPFS distributed storage system and smart contract technology, where the content generated by users on the platform that users consider valuable and need to be confirmed copyrights is presented in the form of NFT. Allowing users to use community tokens to trade with each other to encourage content originators. The specific process is as follows.

The user logs into their account and publishes UGC content, depending on the UGC content they can choose whether the content requires to be copyrighted. When the user chooses not to require it, the UGC content is set to Public, i.e. publicly displayed for viewing; when the user chooses to require it, the next step is to make the UGC content into NFT. After entering the process of making NFT, the system automatically extracts keywords from the UGC content, configures the corresponding JSON file according to the keywords, and then packages and uploads it to the IPFS network. IPFS network feeds the corresponding CID and URL, and the

CID and URL will be the unique identifier of this UGC content in the decentralized network flow. Then is followed by the minting NFT of UGC content based on the ERC721 smart contract. The NFT is a data structure that consists of a JSON file for metadata, and user addresses return a unique digital equity certificate to the creator upon successful minting and are automatically published to the decentralized platform. When other users find the content has purchase value or is collectible, they will need to pay a certain amount of platform tokens to obtain the digital equity certificate. The NFT process is as follows.

User → Submit(require)

Key(UGC)+UGC→upload(in IPFS)

IPFS→Return(URL, CID)

ERC-721 → Mint($\mathrm{Meta}_{\mathrm{json}}$, $\mathrm{Address}_{\mathrm{User}}$, $\mathrm{Token}_{\mathrm{id}}$)

Return(NFT) → User

The figure 4 illustrates utilization of NFT technology in a decentralized UGC contents platform.

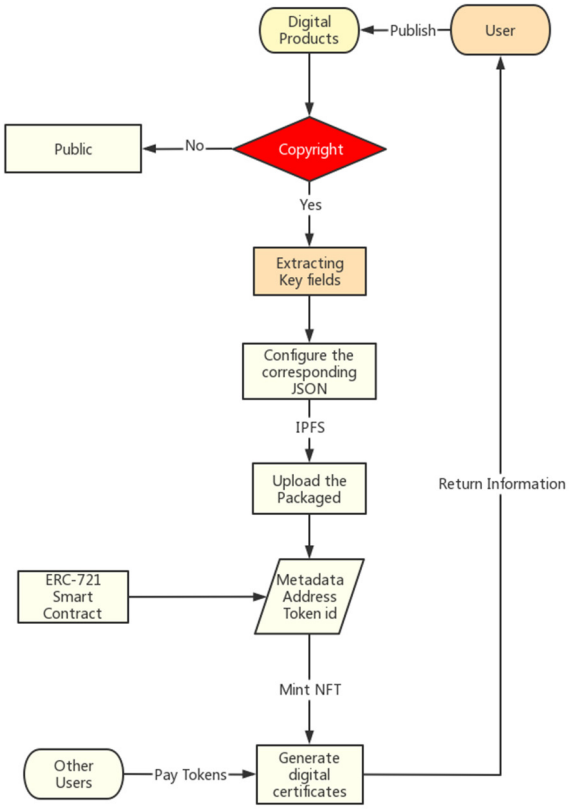

Figure 4 Decentralized UGC publishing flow chart

4 Conclusion

Throughout the whole course of media and social progress, new technologies often gain promotion and development by solving existing problems, and at the same time, they also reshape the entire industry ecology. The development of information technology and the reduction of the threshold for content production and upload have jointly brought about the emergence of UGC mode, but the problems that have come with it also restrict the development of UGC mode. The traceability, decentralization and other characteristics of blockchain and NFT technology provide a new solution for the problem and also paint a new picture for the future internet content production mode. However, technology cannot solve all problems. The introduction of super node voting for verifying information also inevitably centralizes again, giving mainstream media greater authority. Furthermore, there are still large gaps in the docking of NFT rights and existing laws, which not only require the development and improvement of technology, but also require the efforts of the legal profession and the whole society.

References

[1] Valcke P, Lenaerts M. Who's author, editor and publisher in user-generated content? Applying traditional media concepts to UGC providers[J]. International Review of Law Computers & Technology, 24(1) 119-131. doi:10.1080/13600861003644533

[2] OECD.Participative web and user-created content (UCC): New report from organization for economic co-operationand development [OL]. [2008-10-23]. http://www. biac. org/members/iccp/mtg/2008-06-seoul-min/ 9307031E. pdf.

[3] Wen Zongchuan,Chen Qingyu. Research on the collaborative development of subject elements of UGC platform based on the perspective of value co-creation [J]. Modern audiovisual,2022(08):55-59.

[4] Duan, X. and Dai, L. Study on the Relationship between Experience Quality, Co-Created Value and Users' Behavioral Intention of Making Continuous Contributions on Internet UGC Platforms, Journal of Service Science and Management 11 (2018) 267-277, doi: 10.4236/jssm.2018.112019.

[5] CNNIC. Statistical report on the development of Internet in China. [EB/OL] (2021-2-23)http://www3.cnnic.cn/n4/2022/0914/c88-10226.html

[6] M Tambuscio et al. Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web (May.-2015), 977–982.

[7] Lili Xin, Bo Zhang. Application of Blockchain News Production Based on Digital Encryption Technology[C].2020 International Conference on Computer Engineering and Application (ICCEA).

[8] Zhou Maojun,Du Yongxin. Rumor propagation and collaborative governance mechanism in major public emergencies--a study based on rumor information during the New Crown Pneumonia epidemic[J]. Media Watch,2020(09):15-23.

[9] Huang Jie, Zhu Yue. Beijing Internet Court's first case opens [N]. Legal Daily,2018-10-31(003).

[10] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system.(2008)[J]. 2008.

[11] Marco Iansiti and Karim R Lakhani. "The Truth about Blockchain" (2017) 95 Harvard Business Review 118.

[12] Casino F, Dasaklis T K, Patsakis C. A systematic literature review of blockchain-based applications: current status, classification and open issues[J]. Telematics and Informatics, 2019, 36: 55-81.
[13] Zhang Yue. Research on user influence analysis techniques in social networks [D]. Harbin:Harbin Institute of Technology Press,2015.
[14] Liu Rong-San. Research on social network event hotness prediction [D]. Harbin:Harbin Institute of Technology Press,2017.
[15] Esther Thorpe ,Zhang Jianzhong. Revenue diversification: NFT, e-commerce and subscriptions [J]. Young Journalists,2022(01):96-97.DOI:10.15997/j.cnki.qnjz.2022.01.022.
[16] Zhang Sen, Jin Yexian. Opportunities and risks facing journalism in the context of NFT technology application [J]. Young Journalists,2022(15):68-70.DOI:10.15997/j.cnki.qnjz.2022.15.015.
[17] Ifeanyi E Okonkwo, NFT, copyright and intellectual property commercialization, International Journal of Law and Information Technology, Volume 29, Issue 4, Winter 2021, Pages 296–304, https://doi.org/10.1093/ijlit/eaab010
[18] Randhir Kumar,Rakesh Tripathi.Implementation of Distributed File Storage and Access Framework using IPFS and Blockchain [C].2019 Fifth International Conference on Image Information Processing.