Research on Pesticide Registration Test and Traceability System Based on Blockchain

Yuhang Ding[1,a],Fei Sun[2,b] Jun Zhao[3,c],Shenxin Liu[4,d],Zifeng Wang[5,e],Jie Huang[6,f]

{[a]878462161@qq.com,[b] sunfei@hfuu.edu.cn,[c]2430366674@qq.com, [d]3020559376@qq.com, [e]2293255057@qq.com, [f]1373953675@qq.com}

School of Artificial Intelligence and Big Data, Hefei University, Hefei, 230601,China[1,2,3,4,5]
Anhui Chaoqing Technology Co.,ltd,Hefei, 231600,China[6]
Correspondence: sunfei@hfuu.edu.cn, School of Artificial Intelligence and Big Data, Hefei University, Hefei, 230601, China[2]

Abstract. In order to solve the problems of the lack of information management of the detection data in the traditional pesticide registration test, the detection data is easy to be tampered with and cannot be effectively traced,and the centralized storage data is vulnerable to internal and external attacks.In this paper, the technology of combining blockchain and traceability model is used to link the multi-source heterogeneous data after Kettle fusion through traceability PROV modeling,thus solving the problem of data tampering in the detection process and realizing the reliable traceability of data in the pesticide detection process.Therefore, this research has realized the data in the pesticide registration test can not be tampered with and that the reliable traceability of the source and behavior of the data,ensures data security, improved the in-depth development of the pesticide registration data and expanded the traceability ability of the traditional blockchain.

KeyWords: Provenance,PROV,Blockchain,Pesticide detection

1 Introduction

In the field of agricultural pesticide detection, it is vital to use statistical data to analyze the use of pesticides and understand the specific process of pesticide use according to traceability information. However, in pesticide testing institutions, the main task is to generate test reports, and the degree of integration with the new generation of information technology is not enough, and the data development is not enough. A few use big data to conduct quantitative and safety analysis on the creation of pesticides, soil pollution of pesticides, pesticide spraying[1] and pesticide residues[2], and research on risk early warning model[3].

At present, domestic blockchain traceability technology adopts the cloud environment. For example, Alibaba Cloud blockchain[4], and Huawei Cloud blockchain[4], these on-cloud blockchains adopt cloud-based integrated Ethereum and super ledger. The purpose is to facilitate the rapid implementation of the system and quickly build the service chain.

Many foreign blockchain projects combining traceability technology, such as Provence [6], are blockchain traceability projects based on Ethereum to achieve Dapp. It is mainly used in the

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

DOI 10.4108/eai.27-10-2023.2341907

logistics supply chain to trace the production, manufacturing, transportation and sales of products. The ProvChain system solves the problem of data traceability in the cloud [7].

2 Research contents

2.1 Multi-source heterogeneous data

There will be a lot of heterogeneous data in the process of pesticide registration test, especially in the test process. For example, pictures, documents, videos, audio, etc., this data are tough to be processed into unified structured data in technology due to their inconsistent format. In particular, audio, video and other media data can only be traced through their metadata[8]. At present, the most effective way to deal with multi-source heterogeneous data is ETL. According to different strategies, data extracted from various sources for unified storage. After data cleaning, conversion and loading, the data format is correct and the data is consistent and complete[9].

2.2 Data traceability model

Currently, the main traceability models at home and abroad are OPM, PROV and domestic ProVOC. OPM is the earliest traceability model in design, but its disadvantage is that the concept in design is very vague [10], so the W3C improved and formed the PROV traceability model based on OPM [11].

As shown in Figure 1 below.The PROV model can describe different states and relationships of entities and trace back to the description and behavior of data at various stages [12]. The structure of the PROV model shown in the figure below. Its core is three classes and seven relationships. Among them, Entity represents an entity, Activity represents activity, Agent represents agent and represents seven relationships between each other.

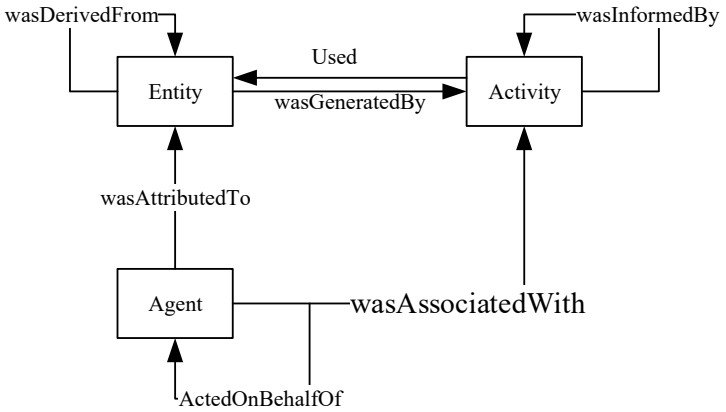

Fig. 1. PROV

2.3 Blockchain platforms

The modular and configurable structure of the super ledger and the pluggable consensus mechanism. In addition, it also has privacy and confidentiality, smart contracts and other features[13]. It has tremendous application potential in finance, government, supply chain logistics and other industries. Therefore, the project uses a super account to establish a blockchain platform.

3 Experimentation

3.1 System design

The project is divided into four parts as shown in the figure. Data collection, data preprocessing, traceability models, and blockchain nodes. The data collection layer is responsible for obtaining data, the data preprocessing layer uses Kettle to ETL multi-source heterogeneous data, the traceability model layer models the fused data, and the application layer divides the entire system into four blockchain nodes. As shown in Figure 2 below.

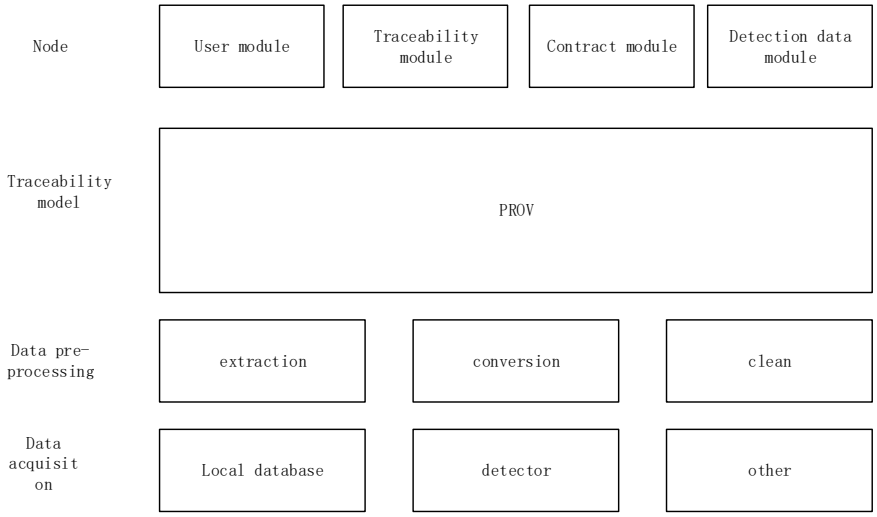

Fig. 2.System architecture diagram

3.2 Multi-source heterogeneous data processing

These data uniformly processed into JSON format through Kettle, which is convenient for each node to upload the corresponding data.As shown in Figure 3 below.

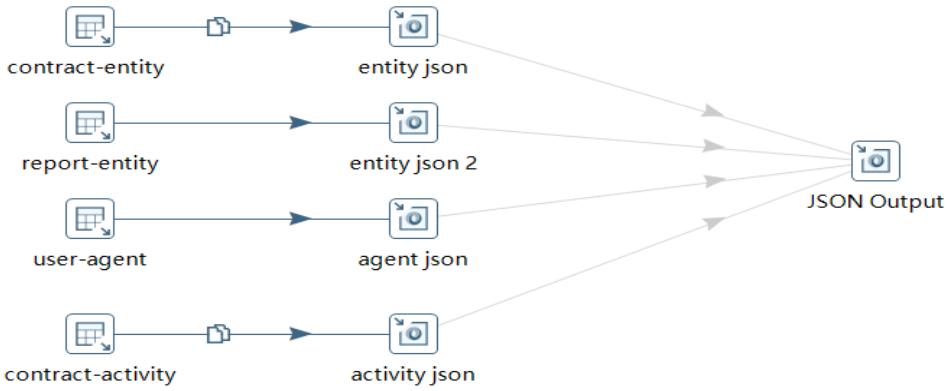

Fig. 3. Construction of traceability model

3.3 Node establishment

Data traceability technology is a method of tracking data, mainly for recording information. It is a record of the metadata of native data, as well as the recording, data flow process, and dynamic changes of derived data.

By establishing a model, on the one hand, it can fully describe the source and behavior of data, and on the other hand, it can reduce the storage of large amounts of data and reduce the time wasted in querying massive amounts of data. Establishing a comprehensive and understandable data model is crucial.

For the whole testing process, there are four parts.signing an agreement, personnel sampling, sample testing, and issuing a report. As shown in Table 1 below,These four parts are involved in the traceability process, that is building blockchain nodes for the user data module, contract management module, traceability module and detection data module respectively.

Table 1. Node mapping table

Structure name	Organization identifier	Organization ID
Peer 1	Org1.example.com	Org1MSP
Peer 2	Org2.example.com	Org2MSP
Peer 3	Org3.example.com	Org3MSP
Peer 4	Org4.example.com	Org4MSP

4 Experimental results and summary

4.1 Experimental result

The container and chain code in Docker are successfully started.After all, nodes started and the chain code installed successfully, you can query the data in CouchDB. Take the detection traceability data as an example,the results are show in the Figure 4 below.

```
"agent": {
        "ag1": {
           "name": "Xiuhua He"
        },
        "ag2": {
           "name": "Xianwen Zhu"
        },
        "ag3": {
           "name": "Kai Zhu"
        },
        "ag4": {
           "name": "Xiuguo Wang"
        }
     },
     "activity": {
        "act1": {
           "type": "object",
           "title": "Hefei-Detection report",
           "prov:startTime": "2020-4-1",
           "prov:endTime": "2021-4-1"
        },
        "act2": {
           "type": "object",
           "title": "Jiangsu-Detection report",
           "prov:startTime": "2020-4-1",
           "prov:endTime": "2021-4-1"
        },
        "act3": {
```


Fig. 4. Experiment results

4.2 Experimental summary

The system uses a traceability model and blockchain to realize data traceability and data security in the process of pesticide registration tests. The function of data tamper-proof and description of data source and behavior is discovered and the data in pesticide registration test can be comprehensively managed and traced. Developed the potential value of test data and expanded the traceability ability of traditional blockchain.

This article mainly consists of three parts. Firstly, using Kettle to process heterogeneous data, and secondly, using the PROV traceability model to model and describe the source and behavior of the data. Finally, it will be combined with the blockchain platform Fabric to store traceability information in Fabric's couchdb.

This system utilizes a combination of traceability models and blockchain to achieve data traceability and security issues during pesticide registration experiments. The function of data tampering resistance and description of data sources and behaviors has been achieved, enabling comprehensive management and traceability of data in pesticide registration experiments. Developed the potential value of experimental data and expanded traditional blockchain traceability capabilities.

But the disadvantage is that the traceability data is metadata, which is the description of the data, and blockchain can only guarantee the authenticity of the data on the chain, and cannot

guarantee the authenticity of the data off the chain. Therefore, in the future, a blockchain based file tamper prevention system can be established to further ensure the security and reliability of data under the chain. This makes the data on and off the chain tamper resistant, and can also be combined with the distributed storage system IFPS to upload and save large files, such as video, audio, and other data that could not have been uploaded to the blockchain.

References

[1] Zeng Xinzhou. Quantitative Analysis of Pesticide Spraying on Pesticide Application Platform Based on Large Data. Journal of Agricultural Mechanization Research.42.09(2020):243-247. doi:10.13427/j.cnki.njyi.2020.09.043.

[2] Chen Zhijun. Safety Evaluation for Pesticide Residue in Vegetables Based on Monitoring Big Data. Chinese Doctoral Dissertations Full-text Database,2019.DOI:10.27630/d.cnki.gznky.2019.000017.

[3] SUN Xiaoming . Study on intelligent early warning model of pesticide application risk based on registration information. Chinese Journal of Pesticide Science. 23.05(2021):1012-1023. doi:10.16801/j.issn.1008-7303.2021.0108.

[4] Blockchain as a Service,2020. URL: https://www.aliyun.com/solution/blockchain/mytc

[5] BCS,2019. URL: https://www.huaweicloud.com/product/bcs.html

[6] Charles Eames.Blockchain:the solution for transparency in product supply chains,2015. URL: https: // www.provenance.org /whitepaper.

[7] LIANGX,SHETTYS,TOSH D,et al.ProvChain: a blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability［C］// 17th IEEE/ACM International Symposium on Cluster,Cloud and Grid Computing(CCGRID).2017: 468-477.

[8] HuiFen Ma.Research and Application of Techniques for Collection and Retrieval on Unstructured Data .Chinese Master's Theses Full-text Database

[9] LIU Chong .Research on integration of college multi-source heterogeneous data. Electronic Design Engineering 23.10(2015):24-26. doi:10.14022/j.cnki.dzsjgc.2015.10.008.

[10] MOREAULL,CLIFFORDB,FREIREJ,et al. The open provenance model core specification (v1.1) ［J］. Future Generation Computer Systems.[2011,27(6)].:743-756.

[11] Luc Moreau,Paolo Missier, PROV-DM:The PROV Data Model,2013.URL: http://www.w3.org/TR/2013/REC-prov-dm-20130430.

[12] Luc Moreau,The PROV-JSON.Serialization,2013.URL:https://www.w3.org/Submission/prov-json.

[13] Yang Cheng, Hyperledger project,2015.URL:https: // www.hyperledger.org.