

ブロックチェーン技術に基づく人文情報活用システムの試作 Prototype of humanities information utilization system based on blockchain technology

磨 有祐実[†] 山田 憲嗣[†] 中山 文[‡] 谷田 純[†]
Ayumi Togi Kenji Yamada Fumi Nakayama Jun Tanida

1. 研究目的

現在、自然科学分野と比べて、人文学分野における各種資料のデジタルアーカイブ化は不十分である。その主な理由として三つが考えられる。第一に、人文学分野における資料には標準化された活字で表現されていないものや欠損があるものが存在し、それらも含めてなるべく元の情報に限りなく近い状態でデジタルアーカイブ化する必要がある[1]。そのため、自然科学分野と比べてデジタルアーカイブ化にコストがかかる。第二に、人文学においては、その分野や範囲の広さから活用される資料の数が多く、しかも増加し続ける。第三に、文献の引用頻度が分散的かつロングテール化し、各資料自体も希少性・複製困難、属人的、所在の分散といった特性を持つため、効率的に網羅することが困難である。これらの理由から、人文学分野におけるすべての資料をデジタルアーカイブ化することは現実的に困難な作業である。そのため、研究者の退職や死去に伴い、後続研究者にとって価値ある多数の資料が廃棄、売却され、散逸している。

そこで本研究では、ブロックチェーン技術を活用したシステムを提案し、その実験システムを試作した。上述した課題に対して、非集権的、かつ、透明性が確保されたブロックチェーン技術を基盤として、たとえ一部であっても、できるだけ多くの資料の基本情報を共有することにより、資料の流動性を高め、それらの有効利用に繋がると考えた。ブロックチェーンは、情報通信ネットワーク上にある端末同士を直接接続して、暗号技術により取引記録を分散的に処理・記録するデータベースの一種であり、単一障害点がないことや、可用性、透明性、改竄耐性の高さなどの特徴を持つ[2]。ブロックチェーン上に記録されたデータは誰でも閲覧できるため、比較的低コストでありながら、広範な利用者をカバーしうる情報活用システムの実現が可能になる。ブロックチェーン技術の応用により、不特定多数の参加者による情報の共有分散管理を効率的に実現できるため、人文情報活用の新たな手法になり得る。本発表では、ブロックチェーン技術に基づくシステム実現における課題を明らかにするとともに、同システムが人文学資料に関わる課題解決に向けてどのように貢献できるかを報告する。

2. 人文情報活用システム

2.1 システム概要

本研究では、研究者が個人で所有する資料の目録を自由に登録・共有しうる実験システムを開発した。ブロックチェーンでは、データの処理や記録に対して、データ量に応じた手数料が必要となるため、資料そのものをブロックチェーン上へ保存することは現実的でない。そこで、実験システムでは取り扱う資料を書籍に限定し、書籍名、著者、出版社、出版年という目録のみを登録・共有することとし

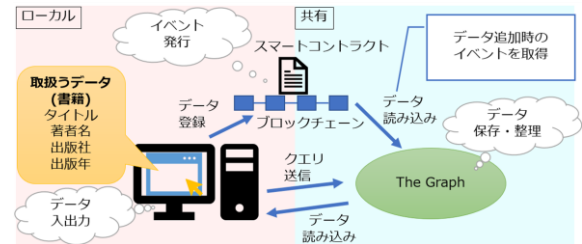


図 1 システムの概念図

た。システムの概念図を図 1 に示す。このシステムではブロックチェーンと The Graph を用いて目録情報を記録・共有する。資料の所有者は、個人のコンピュータでプログラムを起動することで、単一もしくは複数の目録をブロックチェーン上へ送信できる。送信されたデータはスマートコントラクト上で処理され、イベントが発行されることによって The Graph に転送される。The Graph ではデータの整理と保存が行われ、検索が可能なデータベースが作られる。資料を探す利用者は、やはり個人のコンピュータからプログラムを起動することで The Graph に対してクエリを送る。それにより、システムに登録された目録の情報を得ることができ、容易に検索もできる。

ここでスマートコントラクトとは、ブロックチェーン上で動作するプログラムである[3]。スマートコントラクトは条件が満たされると自動的に実行され、その取引内容は誰でも閲覧できる。プログラムがブロックチェーン上にあるため、不正や改竄を防ぐことができ、取引内容やソースコードが公開されるため透明性が高い。ただし、ブロックチェーン上に記録したデータは書き直すことが難しい反面、スマートコントラクトのコードも修正することが難しい。そのため、スマートコントラクトをブロックチェーン上に置いて動作させる際には、エラーやバグがなく、攻撃にも耐えられる状態にする必要がある。

The Graph は、ブロックチェーン上のデータをインデックス化してクエリするための分散型プロトコルである[4]。ブロックチェーンの問題点として、ブロックチェーン上のデータに対して検索やフィルタリングなどの操作が不可能であることや、分散型であるために通信に時間がかかり、比較的簡単な処理でも数時間から数日かかるという特性がある。これに対して The Graph では、ブロックチェーン上のデータにインデックスを付け、データを複数のノードに分散して保存することで、ブロックチェーンの分散性を保ったまま、効率的なクエリを可能にしている。

2.2 データ処理

本実験システムは、Web ベースのフロントエンド、ブロックチェーン上のスマートコントラクト、The Graph の 3 ブロックで構成される。本実験システムは、フロントエン

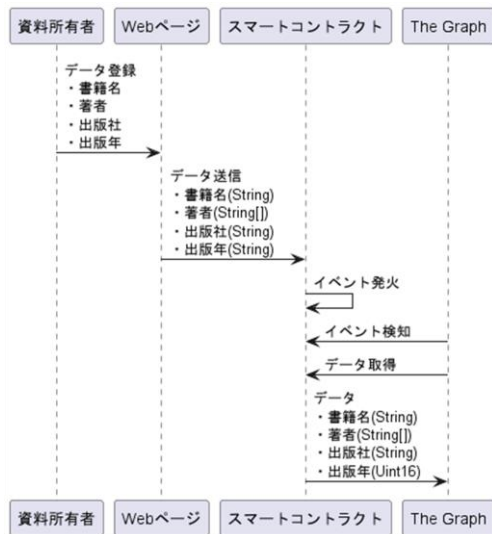


図 2 単体データ登録時のデータの流れ

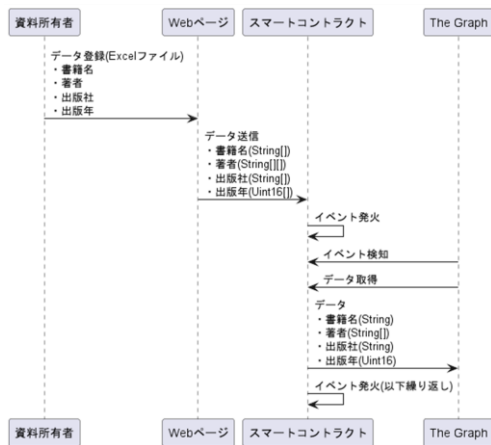


図 3 複数データ登録時のデータの流れ

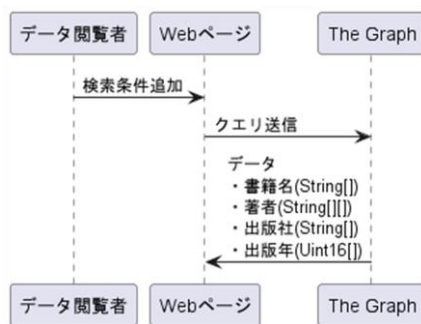


図 4 単体データ登録時のデータの流れ

ドを表示する HTML ファイルと JavaScript ファイルを個人のコンピュータ上で動作させることで、誰でも利用可能である。

データ登録時のデータの流れを図 2、3 に示す。1 件のデータを登録する際、資料所有者によって Web ページのフォームに入力されたデータは、JavaScript のコードによって処理され、スマートコントラクトへ送られる。スマートコントラクトはデータを受け取り、イベントを生成する。The Graph はそのイベントを検知し、どのようなデータが保存されたかを取得し、データをインデクシングして保存する。Excel ファイルからのデータを登録することもでき、その場合は複数のデータをまとめて登録することができる。Excel ファイルを Web ページで読み込むことで得られたデータ群は、JavaScript のコードによって配列としてまとめられ、スマートコントラクトへ送られる。The Graph では全てのイベントに対し、データ取得を行う。

このようにして蓄積されたデータは、図 4 に示す手続きに従って呼び出される。データの閲覧者が Web ページからデータ呼び出しをすると、コンピュータは The Graph の API ヘクエリを送る。この際、閲覧者が指定した著作者名などの検索条件から、JavaScript のコードによって適切なクエリを生成して送信する。The Graph の API からはクエリに応じたデータ群が送信され、データ閲覧者はそのデータを閲覧できる。本来、ブロックチェーン上に記録された大量のデータを呼び出す際には、データを 1 件ずつ呼び出す必要がある。その場合、毎回の通信ごとに中央集権型のネットワークと比較して長時間の通信が必要な上、検索もできない。しかし、The Graph の利用により、これらの問題を解決できる。

本実験システムでは、資料所有者が Web ページから登録したデータは The Graph のみに保存され、ブロックチェーン上にはイベントを発行したという記録のみを残す。The Graph に保存されたデータは永続化がなされており、本実験システム上で検索を行う際には The Graph に保存されたデータを利用するため、他の保存場所を用意する必要はない。

スマートコントラクト活用研究講座

人文情報管理システム

アカウント:

書籍名

著者名

出版社

出版年

excelファイル送信

ファイルを選択

1番目

タイトル: 戏剧肢体语言训练

著作者 1: 玛莉珠

出版社: 开俐实业有限公司出版部

出版年: 1986

2番目

図 5 データ登録用の Web サイト

スマートコントラクト活用研究講座

人情情報管理システム

検索条件を入力してデータを検索し、現在システムに登録されている書籍データがすべて表示されます。
表示するデータを絞り込んだ場合は、検索後に検索条件を入力してください。

※検索条件は必須です。入力内容が空白の場合はエラーを発生します。

検索条件

書籍名

著者名

出版社

出版年

所有者ID

出版年(検索開始年) 年以降

出版年(検索終了年) 年以前

検索条件をクリア

ID	書籍名	著者名	出版社	出版年	所有者ID
0	経済学辞書	丸山修	開明堂出版	1986	
1	新学十年	新学十年編輯小冊、著者	中国出版	1995	
2	新学十年	新学十年編輯小冊、著者	中国出版	1995	
3	新学十年	新学十年編輯小冊、著者	中国出版	1995	
4	経済学辞書	丸山、修	開明堂出版	1986	
5	新学十年	新学十年編輯小冊、著者	中国出版	1995	
6	新学十年	新学十年編輯小冊、著者	中国出版	1995	
7	新学十年	新学十年編輯小冊、著者	中国出版	1995	
8	新学十年	新学十年編輯小冊、著者	中国出版	1995	
9	新学十年	新学十年編輯小冊、著者	中国出版	1995	
10	新学十年	新学十年編輯小冊、著者	中国出版	1995	

図 6 データ閲覧用の Web サイト

2.3 実装

実験システムのフロントエンドは HTML・CSS と JavaScript、バックエンドとなるスマートコントラクトは Solidity で記述した。データ登録時のフロントエンドでは、図 5 に示す入力画面を用意した。実験システムの利用にはウォレットとして、ポピュラーであり操作方法が簡単な MetaMask を採用した[5]。システムの利用時には MetaMask にログインすることでデータの登録ができるようにし、資料データの登録を行った人を特定できるようにした。また、データの閲覧時には MetaMask アカウントは必要ないため、より多くの人がデータにアクセスできる。図 5 の入力ページにおいて、資料所有者がテキストボックスに入力した書籍名や著者名などのデータは、JavaScript のコードで変数に格納される。Solidity でスマートコントラクトに記述した関数を JavaScript で呼び出すことで、変数はブロックチェーンへ送信される。その際には MetaMask から仮想通貨による手数料の支払いを求められ、許可することでデータが送られる。このことは、手数料の確認になるだけでなく、実際にデータが送信されたことの確認にもなる。Excel ファイルを読み込んだ際には図 5 のように、画面上に登録したいデータが表示される。この際、Solidity 上で出版年のデータは Uint16 として扱われるため、数字以外の文字が入っていた場合は警告を表示し、データの読み込みを停止する。その状態でファイル送信を押しても、エラーが発生し、データは送信されない。

スマートコントラクトでは、String 型の書籍名、String 型の著者、String 型の出版社名、Uint16 の出版年の 4 つをひとまとまりのデータとして、1 件のイベントを発行する。Excel によるデータ登録で複数のデータが渡された際には、データの数だけイベントを発行する。つまり、100 件のデータが渡された際には、イベントを 100 回発行する。

データ閲覧時は、図 6 に示すページでデータの閲覧と検索ができるようにした。図 6 のページでは、The Graph に保存されたデータを、登録順である ID 順に並べ、ID、書籍名、著者、出版社、出版年、資料所有者の情報を表示する。このページも HTML・CSS と JavaScript で記述しており、テキストボックスに検索条件が追加された場合、それを含めたクエリを JavaScript で生成し、The Graph の API へ送信する。データ登録時に入力されなかったデータは空欄として表示されるよう設定している。出版年については数字として Uint16 で取り扱っているため、時期による絞り込

みも可能である。また著者については、検索する人名が含まれた全てのデータを表示する。The Graph の API ではデータ数が多い場合、呼び出すデータ数を指定しないと 100 件のみ呼び出される。そのため、データ閲覧時のページではデータを 100 件ずつ表示し、ページを変更するたびに API にクエリを送信することで、データ数に関わらずデータにアクセスできるようにした。

3. 実験結果

本実験システムの開発により得られた結果に対して、人文学における資料取扱いの観点と、ブロックチェーン活用に伴うシステム運用の観点に分けて考察する。

3.1 資料の取扱い

本実験システムは、特定の管理者がない状態で、個人所有の資料の目録を複数の参加者間で共有することができるという点で既存の方式とは異なる。また、資料に登録順の ID を振ることができるという利点もある。ブロックチェーンは時系列データの記録にも適しており、資料の閲覧や貸し出し、譲渡といった資料の動きを記録することで、資料の来歴を公にしなが、発生時期の記録が可能である。

本実験システムでは、誰がどのような資料を所持しているのかを、匿名性を保ったまま公開することができる。また、資料所有者を示すウォレットアドレスは電子メールの送信先として使えるため、資料所有者と資料の閲覧を希望する者との間でコミュニケーションを図ることも可能である。これらの特徴は、個人所有により分散的に管理されている資料の流動性を高めることに繋がる。

3.2 システムの運用

パブリックなブロックチェーンを利用するシステムにおいては、システム利用時に仮想通貨による手数料が必要になる。特に、今回のように同じ作業を繰り返すスマートコントラクトを利用するシステムでは、プログラムの処理量によって手数料が大きく変動する。本報告では、Ethereum テストネットの一つである Goerli ネットを利用した実験結果について示す。

3.2.1 データの保存場所

当初は The Graph だけでなくブロックチェーン上にもデータを保存することを試みた。その際、データをブロックチェーンと The Graph の両方へ保存する場合と、The Graph のみに保存し、ブロックチェーンではイベントの発行履歴のみを保存する場合では、手数料に大きな違いがあった。書籍 1 件分のデータをそれぞれの方法で登録した際の手数料を表 1 に示す。本実験システムでは、何千件という数のデータを保存することが予想されるため、この手数料の差は非常に大きい。そのため、それ以降のデータの保存場所は The Graph のみとした。

表 1 データ保存先と手数料

保存場所	手数料(gETH)
ブロックチェーンと The Graph 2023-04-4-10 8:10:12(UTC)	0.03839117
The Graph のみ 2023-04-10 8:10:36(UTC)	0.01040389

3.2.2 複数データの処理

Excel ファイルからデータを複数登録する際、データを分割して送信する方法と、一括で送信する方法を試した。

分割する方法ではスマートコントラクト側に If 文を入れ、1 件ずつデータの終わりを確認した。一括で送信する方法では、ほぼ同じ内容で If 文が無いプログラムを用いた。2 種類の方法で 520 件のデータの送信を試みた結果を表 2 に示す。これらの実験時には gETH が不足していたため、実際にはデータを送信していない。データを分割して送信する方法では、If 文が 521 回呼び出されており、その処理の分だけ手数料が高くなっていると考えられる。これらの事例から、スマートコントラクトでの処理を減らすことで、手数料を大幅に削減できる。

表 2 データ送信先と手数料

送信方法	手数料(gETH)
分割送信 (2023 年 4 月 5 日 21 時 30 分頃)	0.468
一括送信 (2023 年 4 月 5 日 21 時 30 分頃)	0.34

3.2.3 スマートコントラクト処理

スマートコントラクトの処理はなるべく簡単にする必要がある。なぜなら、ブロックチェーン上に記録されたスマートコントラクトの修正は難しく、またスマートコントラクトでの処理やデータ保存は、それに応じた手数料を要する。特に今回のシステムでは登録されるデータの件数が増えることが想定され、システムの安定動作・手数料削減のためにも、フロント側で処理を済ませることが理想的である。フロント側で行うべき処理には、登録内容の形式の確認や、登録済みのデータとの重複確認などが該当する。

4. 検討課題

現状では、重複するデータをチェックする機構が存在しないため、重複データの登録が可能である。また、実際には存在しない、もしくは所有していない資料のデータ登録も可能である。これらはシステムの信頼性を損なう脆弱性になる。ブロックチェーンの利用により、一度登録されたデータが改竄されないことは保証されるが、登録時のデータの信頼性を確保する方法は、今後の検討課題である。

本実験システムはパブリックなテストネットを用いて実験を行っているが、ブロックチェーンにはプライベート型やコンソーシアム型がある。これらは管理者を必要とする代わりに、ブロックチェーンの利用手数料が変動してしまう金銭的な不安定さが解消される。本実験システムでは、関連研究者の利用を前提としているため、参加者を限定でき、金銭的な問題も緩和されるという点においてプライベート型やコンソーシアム型のブロックチェーンの方が有利である。しかし、非集権的、かつ、透明性が確保されるブロックチェーンの特性はパブリックな環境においてこそ有効に機能する。今後、どの形態のブロックチェーンが適しているか、検討を重ねていく必要がある。

この問題に対して、システムの安定運用は一つの視点を与えてくれる。従来、本実験システムと同様の機能を持つシステムは、単一もしくは複数に冗長化されたサーバーによって実現されてきた。従来のサーバーを用いたシステムでは、システムの利用開始だけでなく、利用の継続にも費用が発生し、管理担当者や責任者も必要となる。また、サ

ーバーの不具合やメンテナンスにより、システムの停止期間が生じる。それに対して本実験システムでは、費用が生じるタイミングは、プログラムをブロックチェーン上に登録する時と、データを登録する時のみであり、それ以外の金銭的な管理コストは存在しない。また、ブロックチェーンは実質的なゼロダウンタイムを実現しているため、可用性が高い。

また、個人が運営する Web サイトと比べると、本実験システムではより容易に参加者を増やすことができ、管理の負担も分散できる。個人の Web サイトで同様の機能を実現する場合、資料所有者と Web サイト運営者との間で情報のやり取りを要し、時間と労力が必要となる。それに対して本実験システムでは、資料所有者がデータの登録を行うため、データ管理者と資料管理者を一致させることができ、システムの運営を行う管理者に負担が偏ることがない。このことは、参加者の増加を容易にするため登録データ数の増加に繋がり、運営のための負担を分散できることから継続的なシステム運用が期待される。

さらに、広く運用されている機関リポジトリと比較した場合、資料の内容へアクセスすることが難しい反面、資金や人員が不足しているような研究者であっても、他の研究者の資料にアクセスしたり、自身の資料を提供したりできる。また、デジタル化が難しい資料について除外することなく扱える。

5. まとめ

本研究では、各種資料のデジタルアーカイブ化が不十分な人文分野における課題を解決するため、ブロックチェーン技術を活用した人文情報活用システムの実験システムを試作した。このシステムではブロックチェーンと The Graph を用いて目録情報を記録・共有する。パブリックな Ethereum テストネットの一つである Goerli ネットにおける開発システムの動作を確認した。これにより、非集権的、かつ、透明性が確保されたブロックチェーン上での人文情報共有システムの運用にかかる課題を明らかにした。

現状では、重複データへの対応は未実装であるが、それ以上に登録データの信頼性を確保する方法の検討が必要である。また、ブロックチェーンの三つの運用形態のうち、人文情報活用システムにおいて、いずれの運用形態が適しているかは重要な課題となる。システムの安定運用や広範な利用者へのサービス提供という観点ではパブリック型ブロックチェーンが優れており、今後、実用システムを目指した検討・設計を進めていく。

参考文献

- [1] 八村 広三郎, “人文・芸術系のデータベース-今そしてこれから- : 1. 人文科学とデータベース”, 情報処理, 38, 5 (1997).
- [2] Nakamoto, Satoshi. “Bitcoin: A peer-to-peer electronic cash system.” Decentralized business review (2008):21260.
- [3] 山崎 重一郎, 安土 茂亨, 金子 雄介, 長田 繁幸 “ブロックチェーン技術概論 理論と実践”, KS 情報科学専門書 (2021).
- [4] The Graph について, 入手先 <<https://thegraph.com/docs/ja/about/>> (2023.04.10).
- [5] MetaMask, 入手先 <<https://metamask.io/>> (2023.04.10).

† 大阪大学大学院情報科学研究科 Osaka University, Graduate School of Information Science and Technology

‡ 神戸学院大学人文学部 Kobe Gakuin University, Faculty of Humanities and Sciences