

第8班 共同研究

『非文字資料研究のコミュニティにおける 知識とサービスの効率的な検索と安全安心な流通研究』

木下宏揚 佐野賢治 能登正人

森住哲也 宮田純子 小松大介

KINOSHITA Hirotsugu SANO Kenji NOTO Masato

MORIZUMI Tetsuya MIYATA Sumiko KOMATSU Daisuke

1 研究目的

本共同研究は、非文字資料を研究者間および専門家以外の人との間で情報の提供、共有などを行うために必要な基盤技術を構築し、実際の資料や研究者などを対象とした実証システムにより、その有効性を検証することを目的とする。第3期共同研究計画における非文字資料の検索、流通等に関する成果を踏まえて、第4期共同研究では、インターネットエコミュージアムや只見町に開設予定の民俗博物館において必要なデータマイニングやデータの入力や検索に適したユーザインタフェースなどの基盤技術を開発することを目的とする。具体的な事業内容は以下の項目からなる。

1. 知識とサービス、物の流通と価値交換

非文字資料を研究者間および一般ユーザと知識、サービス、資料をやり取りするためにゲーム理論によりモデル化を行う。また、ビットコインで注目されている自律分散的に事象や価値の移転を記録するブロックチェーンの技術を用いて安全安心な価値交換を行うシステムを構築する。

2. 知識とサービスの検索とマイニング

非文字資料のデータベースや研究者が資料を検索する際に、作業の流れであるコンテキストの一面に着目して情報の類似度などに基づいてファイルの自己組織化などによりユーザに対して最適な情報を提示する。

3. 個人情報や重要情報、著作権の管理

ブロックチェーンを用いたアクセス履歴管理を行うことで個人情報に対するハイパーグラフを用いた推論攻撃検出の実装を容易にする。また、ブロックチェーンによる価値移転により著作権管理を効率的に行う方法を提案する。

2 活動経過

2.1 2017 年度研究経過

2.1.1 ブロックチェーンを応用した流通と著作権

デジタル化された著作物がオリジナルであるという証明や二次著作物、不正使用の対策には埋め込みの順序の証明に意味を持つが電子透かしのみでの実現は困難である。先行研究では、信頼できる第三者により解決しているが、プライバシー、結託、第三者の存続性の問題があるため完全に信頼することが難しい。信頼できる第三者を必要としない方法を提案するために、本研究では、ビットコインなどに用いられているブロックチェーンの仕組みを電子透かしに用いる方法について考察する。ブロックチェーンは、記録する情報をハッシュ関数により圧縮するが、画像など冗長性の多いデータは、単にハッシュ関数を求めただけでは不可逆な圧縮符号化や電子透かしの埋め込みなどにより同一の画像にもかかわらず異なるハッシュ値となってしまう。そこで画像の特徴量に基づくフィンガープリントを生成し、これのハッシュ値を用いることで電子透かしの埋め込みにも対応可能となる。フィンガープリントを生成する手法として、python の anaconda 環境において著作物を Bag-of-words 表現の SIFT を用いて登録する。画像一枚の特徴を抽出し得られた特徴量からハッシュ値を生成し、このハッシュ値をブロックチェーンに登録するという方法を採用する。今年度は画像の特徴量抽出について検討を行った。

2.1.2 大規模クラウドに於ける新しい情報セキュリティを潜在的な非文字概念から探究

非文字として表象されるテキストとは何か？この問いに対し、今年度は下記に示す主要な成果（論文3点、他）が得られた。即ちまず、「非文字」とはテキストに陽に書かれていない潜在的なテキストであると見做す。その潜在性を工学的に示す試みの1つとして、AI で使用される確率的モデルに着目する。非文字に於ける潜在性は潜在的ディリクレ配分法に於ける潜在的パラメータと呼ばれる確率変数によって示される。本論文では次に示す4つの仮説に基づいてこの問題に取り組んだ：

1. 『意味』は『作用』とともに変動する。
2. 思考と言語的振る舞いには情報理論的な相関がある。
3. 実在を指示するテキストの、「語」を命題の変項と見做し、「語」という変項を『確率変数として見る』。
4. テキストを構成する語は事前的な語と事後的な語の相互情報量が最大になる様に生成される。

潜在的なデータを顕在化させる取り組みの具体化の1つとして非文字データの分類整理のみならず、情報セキュリティに於けるセキュリティモデルの根本的な見直しにも取り組んだ。即ち隠れチャネルモデルと推論攻撃モデルの統合である。この取り組みは非文字とは無関係である様に見えるが、根本は類似した現象である。重要な手掛りは、ウィトゲンシュタインの『家族的類似概念』、及び、『私的言語の捉え方』である。「テキストのあつまり」とは、＜私＞と他者の言語的振る舞いが映し出された object であり、かつ＜私＞と他者の「その時々」が示された「対象領域」：『ウィトゲンシュタインの言語ゲームに類する帰結』である。次にこの「対象領域」の分析手法の必要条件としてテキストが生成、結合、消滅する過程を確率測度空間の写像、即ち確率変数で表現し、潜在的ディリクレ配分法に基づく確率的モデルで記述する方法を提案した。さてこの様にして探究してきた確率的モデルは論理学的モデルと如何なる関係にあるのだろうか？私たちは日常的存在者を如何に捉え様々な規則に対峙するべきなのだろうか？言い換えれば、「世界を確率的に捉える事」は、倫理として如何に捉えられるべきであろうか？今季は、この問いに対する試論にまでたどり着いた。即ち：＜私＞的言語の

パラドクスを認めながらも、実はその治療法たる“文法像”と言う概念が＜私＞そのものを示すパターンであり、そのパターンを示す必要条件が確率測度空間であること、確率測度空間における倫理とは、確率的存在者に嵌める制約、或いは「枠」であり、自他それぞれにとっての振る舞いの調和が映し出された“文法像”の1つの表現形式であることである。

2.1.3 推論攻撃を考慮した重要情報漏えい保護

一般に推論攻撃によってどのような情報が得られるかは明白ではないためデータベース管理者は、扱う機密情報に対して推論攻撃が成功する可能性をあらかじめ考慮して情報を管理しなければならない。そのためには、データベース管理者（情報発信者）が機密情報間の依存関係をあらかじめ把握しておくことが重要だと考える。先行研究では、このような推論攻撃に対抗するためには膨大な情報群とその間にオブジェクト間の推論関係を検知して警告するようなシステムが必須であると述べた上で、グラフにリスト着色を施すことで推論経路を検出し、アクセス制御をする手法が提案されているが、そもそものオブジェクト間に存在する推論関係というものはあらかじめ人間が定義した上で推論経路を検出している。本研究では、昨今の扱うデータの多様性に対応すべく、対象とするデータを単語レベルから文書レベルにまで拡張し、推論規則を求めるために必要となるオブジェクト間（情報間）の関連性を確率的に求める手法としてトピックモデル（潜在的ディリクレ配分法）を用いた。これにより、情報漏洩を引き起こす危険な投稿を未然に防ぐ事ができる。

2.1.4 コンテキストに基づくファイル管理

トピックモデルは潜在意味解析などに用いられていることから、検索シーンにおいてユーザが潜在的データの検索が可能だと考える。先行研究ではトピックモデルの画像分類への適用や画像のトピックを抽出し、トピック分布により類似画像を検索する方法が提案されている。文献では画像の特徴量である SIFT を抽出しベクトル量子化を Bag-of-words 表現（BOW 表現）できる形式にし、k 平均法でクラスタリングしてトピックモデルを用いて画像を分類している。画像だけでなく文書の情報を加えることにより、画像もしくは文書データからのみでは発見することが困難な潜在的な画像発見が可能となる。本研究では、文書と画像を BOW 表現しトピックモデルに適応させる。画像においては SIFT により特徴量を抽出し、ベクトル量子化を用いることで BOW 表現できる形へ変換する。文書と画像から BOW 表現により統合し単語集合を抽出し、これに対してトピックモデルを適用し、潜在的データを検索するためのフレームワークの提案を行う。これにより Wikipedia など文書が画像を説明しているようなウェブサイトなどから関連または潜在的な情報を抽出することでユーザに新たな発見を促すことができる。

2.1.5 テキストの相互情報量により非文字オントロジー間を接続する概念の提案とそのケーススタディ

研究の目的に鑑み、最も根本的な問題：「非文字を分析可能なテキストとして如何に捉え処理するか」、と言う問題に対し、確率論的解決の試みを示す。即ち、非文字オントロジーを含む大量の日常的テキストを想定し、そこから語と語の間の局所的で相対的な自己相互情報量により語の連鎖を生成する。そして語の連鎖が生成するテキストを情報理論的な相互情報量として評価する手順を示す。非

文字データベースの検索はこの手順で生成される大規模なグラフの検索問題に帰着される。本論文では、異なる木構造の間をテキストの集合に於ける語と語の自己相互情報量と言う内積の連鎖で接続し、生成した語の経路が生成される事後確率密度、及び経路から潜在的テキストが生成される事後確率密度を潜在的ディリクレ分布配置法 (LDA) でベイズ推定する方法を示す。更に、自己相互情報量の計算によって異なるオントロジー語間を連鎖させて接続するアルゴリズムについて、自己相互情報量を計算する関数を再帰的に使用するアルゴリズムを提案する。また、異なるオントロジー語間の経路作成が可能であることを case study によって示す。

2.2 2018 年度研究経過

2.2.1 自己組織化ファイルシステム

作業のコンテキストに応じたファイルのクラスタリングとディスプレイへのアイコン配置の最適化の手法として、群知能のうち PSO を用い、目的関数に Boid 的要素を用いるモデルの提案を行った。

2.2.2 Topic Model とオントロジーを用いた非文字検索システム

先行研究では福島県只見町の民具を対象とした民具オントロジーを考慮したデータベース (以下、DB) 検索方法が提案されている。しかし先行研究の民具オントロジー (以下、既存オントロジー) では概念系の定義が不透明なため、検索できない民具がある。一方で、農作業概念を記した語彙体系である Agriculture activity ontology (以下、AAO) を用いることで、正確に情報の意味を考慮し、結果を推論する意味的検索を実現できる可能性がある。本研究では、生産用途民具を対象とし、AAO ベースの民具オントロジー (以下、民具 AAO) の構築、それに対応した DB 構造を提案し、高度な意味的検索実現を目的とした。現状の研究では民具という「扱い難い性質」をオントロジーという木構造の分類整理手法によって表現する、その手法自体に限界がある。本研究では非文字の本質を情報理論的に捉える。言い換えれば、木構造の分類整理法ではなく、民具などの対象物の分類をクラスター分析として確率論的に捉える。対象物のクラスターを確率分布として捉えることにより、この問題は Bayesian 機械学習の問題に還元される。この研究では非文字の意味を確率論的な潜在性として再定義し、対象物のクラスターの中の潜在的な非文字の確率変数によって木構造で表現される概念をつなぐモデルを示し、それを実現するシステムを提案した。

2.2.3 ハイパーグラフを用いた推論攻撃検出

情報漏えいの原因として注目されている推論攻撃をハイパーグラフを用いて解析する手法を開発してきたが、推論規則の獲得手法について検討を行った。ベイジアンモデルによる情報漏えい分析のための機械学習では、確率測度空間上のベイズ主義的な事象と見做す。即ち、機密にすべき事前的、潜在的なクラスターが機密を包摂するクラスターの族と言う事後的なクラスターを形成すると解釈し、潜在的な確率変数を推定する。現象をこの様に捉えるベイジアンモデルにより、観測されたテキストデータを条件とする covert channel と inference channel の潜在的な確率変数の事後確率を統一的に機械学習するモデルを提案した。更に提案モデルの事後確率は word2vec により機械学習可能であることを示した。

さらに、自己相互情報量 (PMI) をベースに PMMI(Pointwise Mutal Multiple Information) という指標を用いた評価方法の提案を行い、統計的に分析をすることで機密情報に対する推論規則生成に適用した。次に、米国や英国をはじめとした世界各国で標準化されたデータ形式である Linked OpenData (LOD) に基づいたデータ公開が行われている。LOD はデータのオープン化、分野内でのデータ共有、そして分野を横断したデータの共有を促進するという特徴を持っている。本研究では、Ubuntu、OpenRefine、Google ドライブの環境下で LOD を実装し、コミュニティでの LOD と公共の LOD のリンクによって起こる Covert Channel を分析し、情報漏洩防止を目的とした。OpenRefine の Reconciliation Service API (名寄せ) アルゴリズムの中にコミュニティでの LOD と公共の LOD の間の Covert Channel を分析し、アクセスを制御するアルゴリズムを考察した。

2.2.4 ブロックチェーンを用いた推論攻撃の検出法

情報漏えいが発生した際、情報の持ち主が自分の情報が漏えいした事実を知らないため、詐欺などに遭う可能性が高い。ブロックチェーンはアクセス履歴の信ぴょう性を保証してくれるため、それに記録されたアクセス履歴に基づいて情報が漏えいする可能性があるかどうかを確認できるだけでなく、ユーザに自分の情報が漏えいするかもしれないということを警告することによって詐欺などからユーザを守ることにもできる。情報漏えいが発生する原因としては、アクセス権の矛盾に起因する Covert Channel がある。従来はアクセス権の設定が固定された静的な Covert Channel 解析を行っていたが、アクセス権の変化をブロックチェーンに記録し、動的な Covert Channel 解析を行うことで静的な解析では検出できなかった情報漏えいの検出を可能にした。通常のオンラインストレージは管理者がユーザごとにアクセス権を分離して割り当てているため情報漏えいは発生しないが、共同研究の情報共有に必要不可欠なオンラインストレージの招待機能はファイルに対して複数のユーザのアクセスを許可するため意図しない情報漏えいが起こる。招待によって起こる第三者への情報漏えいを防ぎ、アクセス権の変更をスマートコントラクトを用いて自動化を目的とするシステムを提案した。スマートコントラクトの特徴である決めた条件を満たしたら、契約内容を自動的に実行することを利用して情報漏洩を防ぐためにアクセス権を自動的に変更できる。また自動的にアクセス権が変更できるため、管理者が必要なく、ユーザ間で自由に招待でき、安全に情報共有ができる。

2.2.5 ブロックチェーンに基づくデジタル画像の著作権管理システム

オリジナルの著作物をもとに、新たな著作物を創作する二次著作物、三次著作物に電子透かしを埋め込む場合、著作物が作成され新たに埋め込まれた電子透かしの順序関係を明確にする必要がある。従来は信頼できる第三者が透かし埋め込みの順序を保証していたが、自律分散型のシステムでは望ましくない。そこで、透かし埋め込みの順序をブロックチェーンに記録することで、これを解決するシステムを提案した。また、透かし情報を別の画像に流用できないように、人間が同じ情報と感じれば同じメッセージダイジェストを生成する知覚ハッシュ関数に基づいた透かし情報を利用している。また、セキュリティモデルに Take-Grant モデルを採用することで、デジタルコンテンツのやり取りにおいて利用権などの改ざんがされていないかを確認することを目的とした。

2.2.6 ファイアウォールのポリシー設定問題

ファイアウォールのポリシー設定問題は、先行研究では機械学習のナイーブベイズを使用して、確率的性質からアクセスコントロールリスト (ACL) を更新する手法が試みられている。しかし確率を使用するため、ポリシーに従わない異常なルールが作成される場合がある。そこで教師あり機械学習のランダムフォレストを用いて、ポリシーを満たしたルールを自動で ACL に適用することができる。と考える。よって本研究ではランダムフォレストを用いてポリシーを満たすルールの作成と ACL への自動適用を目的とするシステムの提案を行った。また、このファイアウォールを仮想環境化するにはどのようにすればよいかも提案した。

2.2.7 画像処理に影響を受けない特徴量の抽出の基礎的研究

Deep Learning が注目された大きな理由は識別に有効な特徴量を機械が自ら学習することにある。従来の機械学習では識別に有効な特徴量を選択する必要がある。その先駆けとなる手法の一種として Convolutional Neural Network (CNN) がある。CNN は特に画像識別に特化していて、識別精度が高いともいわれている。その識別精度が高い理由としては中間層に畳み込み層とプーリング層という 2 種類の層を用いていることである。しかし、中間層に着目した研究が少ない。そこで CNN の中間層を抽出しその特徴量を報酬とみなし学習させるシステムを検討した。サンプル画像に対してアフィン変換、ノイズの付加など様々な加工をした画像データを CNN の実装である Caffe や Keras を用いて中間層を抽出する。その抽出した中間層の特徴量を報酬とみなして、ベイジアン逆強化学習で学習させ、そこで不変量が得られるかどうか調べるためのシステムを検討した。

2.2.8 只見町の調査、研究打ち合わせ

9 月 19 日から 20 日にかけて只見町に調査、研究打ち合わせの出張を行った。旧朝日公民館で民具資料の展示方法、分類方法、聞き取り方法の調査と主な民具の製造過程や使用方法について聞き取りを行った。只見町役場において現状報告などの打ち合わせを行った。また町長、副町長を交えて情報交換を行った。只見町役場において只見カードの調査と今後の方針について打ち合わせを行った。数年後に新設される只見町の博物館において研究成果を取り入れたシステムを設置できればということになった。只見ブナセンターにおいて新国氏より展示方法や資料収集方法について説明を受けた。

2.2.9 潜在的なテキストのパターンを AI により健在化し、分類評価する研究 (パターンランゲージの研究)

(1) ノンパラメトリックベイズによる「『多重な』潜在的ディリクレ過程」を詳細設計した。具体的には PMI の評価指標によって生成する単語の連鎖によって 2 つのオントロジー語彙をディリクレ過程により接続するモデルを学会発表 (9 月) した。

(2) パターンランゲージとは、言い換えれば「非文字を確率測度空間から測度空間へ写像する確率変数である」と捉える言語である。或いは、パターンランゲージは大量のテキストの中に潜む潜在的な確率変数である、とも言える。潜在的な確率変数は既知ではないので、まず機械学習させるテキストを効率よく収集するシステムが必要である。本年度は単語間の PMI により単語の連鎖を生成し、単語の連鎖の KL 情報量により単語連鎖の Boid (但し、各単語はそれらが帰属するテキストに紐付け

られていなければならない)を生成する手法を提案した(卒研テーマ)。

(3) パターンランゲージは自他のための新しい言語である。それはウィトゲンシュタインの家族的類似概念によって言語ゲームを実践する場所を提供しなければならない。そのためには大量のテキストの潜在的なパラメータを確率変数とし、上記(2)の処理の後、テキストのクラスターの確率分布を機械学習させる。今年度はこのモデルとして Blei の supervised LDA の潜在的確率変数を強化学習のアクションの潜在的なパラメータと解釈し、かつ Ramage の Labeled LDA を強化学習の状態 S のラベルとして解釈する、教師ありベイジアン逆強化学習 (Supervised Bayesian Inverse Reinforcement Learning (S-BIRL) と呼ぶ事にする) のグラフィカルモデルを提案した (3月学会発表)。

(4) 上記に示すモデルは<私>という視点を人工知能的システムに組み込む「離見の見」の概念を実現するためのアプローチである。即ちそれは、「それぞれの<私>が他者をいかに解釈するか」という問いに対し、ウィトゲンシュタインの言語ゲームに於ける家族的類似の概念と世阿弥の「離見の見」の概念を確率モデルによって設計するという位置付けにある。この研究では、そのような設計が、“ただ設計するのではなく、善く設計していることになるのか?”という倫理的考察を、ウィトゲンシュタイン、ベルグソン、マルクスガブリエル、坂部恵、等を手掛かりに考察を継続し、設計に反映させた(学会発表)。

2.3 2019 年度研究経過

2.3.1 ブロックチェーンの電子透かしへの応用

ブロックチェーンは信頼できる第三者に依存することなく、権利の移転などのイベント発生の時系列の保証を行うことができる。ブロックチェーン技術に基づいた仮想通貨やスマートコントラクトが普及してきており、著作権管理の分野にも利用され始めている。これを電子透かしを含むデジタルコンテンツの著作権管理に応用した。

ブロックチェーンによる電子透かし管理システムの改善と透かし情報の削減

昨年度提案したブロックチェーンを用いた多重電子透かし管理システムの、電子透かし、ブロックチェーン、および知覚ハッシュに基づくデジタル著作権管理システムの3つの相互関係について検討を行いプロトコルの改良を行った。電子透かしの透かし情報のデータサイズと電子透かしの耐性はトレードオフの関係にある。そこで、ブロックチェーンによる透かし情報の保存と管理の手法において、透かし情報はブロックチェーンに記録し、実際に埋め込む透かしは、その暗号学的ハッシュ値を用いることで、透かし情報のデータサイズを大幅に削減することが可能となった。

新しい知覚ハッシュ関数

信頼できる第三者が不要な多重電子透かし二次著作物など複数の権利者が介在している場合、創作や加工の順序を明示する必要がある。従来は信頼できる第三者が情報を管理していたが、コストやプライバシー保護、セキュリティの観点から望ましくない。ブロックチェーンにコンテンツのメッセージダイジェストを記録することで、これを解決できるが、SHA256 などの暗号学的一方向性ハッシュ関数を冗長性の高い画像情報などに適用すると、加工および符号化により視覚的には差異を検出でき

なくても異なるハッシュ値となってしまう。そこでコンテンツの加工に耐性のある知覚ハッシュが必要となる。深層学習が画像認識などの分野で普及が進んでいるが、畳み込みニューラルネットワークの処理過程で得られる中間層の出力は画像に固有の構造情報が含まれている。そこで、中間層出力から知覚ハッシュに利用可能な最適な組み合わせを検討し、透かし情報に適した知覚ハッシュ構成法を検討する。既に大規模なデータセットで学習済みの CNN を、本実験用に転移学習したものを利用する。転移学習とは、既に大量の画像データセットで画像の分類について学習した CNN のモデルを、別の画像データセットの分類に利用する手法であり、少ないデータセットでも高い分類精度が期待できる。また、CNN はフィルタを通して画像の特徴を抽出する畳み込み層と、特徴毎にさらに小さな画像を生成していくプーリング層からなる。プーリング層では、畳み込み層で抽出した特徴を小さな画像にまとめており、浅い層では入力画像とほとんど同じ画像であるが、層が深くなるにつれて、画像がより小さく単純な構造になっていくことから、知覚ハッシュに利用する中間層は最も深いプーリング層を利用する。これに主成分分析を適用し、確率分布を解析することで、加工編集に対して耐性のあるプーリング層のノードを抽出し、これをもとにハッシュ値を導出した。また、CNN 中間層出力データの位相幾何学的構造のクラスター群から不変的なクラスターを抽出し、それをベイズ統計手法によって機械学習推定することにより知覚ハッシュを生成する手法を考案した。

CNN における認識率向上のための層数と Loss 関数の選定

画像認識の発展は、自動運転の実現など産業革命の分野の発展においても必要不可欠である。また、人間をも上回る認識率を誇り、近年大きな注目を集めている。しかし、層数や容量の多い学習済みモデルを利用した画像認識は、時間がかかり効率的とは言いがたい。本研究では、9 層の CNN をベースモデルに利用し、少ない層数による効率的で、現実的な画像認識を目的として、畳み込み層追加と選定、 ϵ の最適値の選定を提案し認識率向上を図った。さらに実験結果を比較検証し、提案手法の有効性を示した。

2.3.2 ブロックチェーンを用いたデジタルコンテンツの流通

デジタルコンテンツとその著作権を保護するため、様々な DRM（デジタル著作権管理）が提案されているが、現在広く普及している DRM では、コンテンツの配信事業者がコンテンツの利用制限を行い、利用者間でのコンテンツの受け渡しは制限されている。これらを改善するため、新たな手法としてブロックチェーンに基づく著作権の移転を可能にするスマートプロパティや、権利に基づくコンテンツに対する操作を保証するスマートコントラクトが DRM に用いられ始めている。従来の DRM を用いて、信頼できる第三者を必要とせず、デジタルコンテンツの著作権や利用者がコンテンツに対する権利を証明することは困難である。これらの問題点を解決するため、先行研究ではスマートプロパティやスマートコントラクトを用いたシステムやビットコインのプロトコルの 1 つである Open Assets Protocol を用いたシステムが提案されている。しかし、ユーザに対してデジタルコンテンツの利用を制限するシステムは提案されているが、デジタルコンテンツが他のデジタルコンテンツの利用を制限することはできない。例えば、音楽配信サービスと楽曲のような関係の場合、コンテンツ利用の制限は音楽配信サービスに依存し、楽曲そのものに著作権は制限をかけることが困難である。よっ

て、コンテンツとその利用者、コンテンツとそれを利用するコンテンツの関係を明確にしてデジタルコンテンツを保護する必要があると考える。提案するシステムを通してデジタルコンテンツを流通させるかぎり、利用者間、コンテンツ間、利用者とコンテンツの3つの関係を合わせた中でコンテンツが保護できるようなシステムを目指す。本研究では、権利の移転を、主体を利用者、対象をデジタルコンテンツとした Take-Grant Model で表現し、不正なコンテンツ利用を Covert channel にモデル化することでコンテンツの著作者が意図しない権利の流通を分析し、流通の防止を可能とすること、これらのシステムをブロックチェーン技術であるスマートコントラクト内で実行することで、利用者は取引実行の有無とコンテンツに対する権利を証明し、安全で効率の良いコンテンツの流通を実現することが目的である。

2.3.3 非文字資料の検索: LDA を用いた非文字資料検索法

近年、インターネット上に日々大量の情報が増えてきている。紙の文書も電子データに変換する試みも増え、書籍を調べるよりもインターネットを利用して調べるほうが素早く多くの情報を比較できる。正確で早い検索を可能にする、文書の意味をメタデータに記述する方法や、文書中の単語を意味解析し自動で分類する方法が研究されている。トピックモデルは文書が複数の潜在的なトピックから確率的に生成しているという考え方で、pLSA や LDA という手法を用いることでコンピュータで文書のトピックモデルを計算することができ、文書のトピック分布を比較することにより文書間の類似度を測れる。そこで LDA の出力にベイズの定理を用いることにより算出される“単語を構成するトピック分布”を用いて単語と文書の類似度を測れるのではないかと予想した。本研究では、LDA の出力にベイズの定理を用いることにより算出される“単語を構成するトピック分布”と文書を構成するトピック分布の類似度から、文書中の特定の単語に注目したときに現れる類似した文書を提示するシステムのモデルを提案した。

2.3.4 重要情報の保護

推論攻撃の情報漏洩に着目した言語ベクトルの次元圧縮

近年では SNS などでの情報発信やビッグデータの解析などによりさまざまな恩恵が受けられる反面、プライバシーの侵害が問題になっている。従来のアクセス制御の枠組みでは扱うことが困難であった推論攻撃による情報漏えいに対処する必要がでてきた。推論規則生成を考えた場合は推移率、相関を扱えるモデルである skip-gram や c-bow 等を用いると高次元データになってしまう。自然言語処理におけるデータの高次元化は精度、計算速度の面において非常に重要な問題である。日本語のドキュメントにおける高次元ベクトルデータに対する次元圧縮の手法において、Tensor 分解を組み合わせることにより低次元に次元圧縮した際の精度の向上を目的とし研究を行った。

非文字データベースを対象とした AHP に基づく covert channel の解析

ストレージからの情報漏洩を防ぐために、アクセス制御が使われている。アクセス行列は一般に covert channel と呼ばれる情報漏洩を引き起こす経路が存在する。従来この covert channel の評価はセキュリティモデルによって行われてきた。しかし、人と人の関係、あるいは人と情報の関係を論理学

的モデルによって示すことは、それによって表現される応用的現場を限定させることになり、使い勝手が悪くなるという問題があった。本研究は神奈川大学非文字資料研究センターで行われている「非文字資料」をデータベース化するという研究に関連している。この非文字データベースにおいても、covert channel が起こり、情報漏洩につながる可能性がある。そこで、大量のテキストをトピック分析し、テキストの確率変数としてのクラスタを学習させ、クラスタの中にあるテキスト同士の類似度を確率的に求める。次に人と人の関係、人と情報の関係をセキュリティモデルの属性から役割、競合、および所有と定義する。最後に、テキストの類似度、役割、競合、所有を階層分析法（AHP）の評価基準と定義し、意思決定を支援する AHP 分析によって複数の covert channel を評価し、切断すべき covert channel を選択する、というモデルを提案した。

2 つの確率モデルの組み合わせによる Multi-Label Learning の解釈について

本研究では、提案したセキュリティモデルに新たに人間の定めたセキュリティポリシーやセキュリティ規則を確実に反映させる手法として Multi-label Learning による重み付けを行うモデルについて提案を行う。前回までのセキュリティモデルでは、人間の定めたセキュリティポリシー等が教師データに入っていたとしてもそれがアクセス制御に反映させることができていたかは、確実性に欠ける点が存在していた。しかし、この重み付けを行うことによりそのセキュリティポリシー等をより確実に反映させることができ、命題および関連するテキストを含むアクセス制御を可能とした。その重み付けとして 2 つの確率モデルを使用し、テンソル分解したものとみなすことができる。これは今までの Multi-label Learning で使用されてきた 1 つの確率モデルによるテンソル分解と同じものであると言える。本研究ではその重み付けの実際にどの程度の効用が得られるかを実験によって確かめた。

ランダムフォレストを用いたファイアーウォールの規則の生成

近年、ネットワークの発達によって技術の向上がみられる反面、これを悪用したサイバー犯罪が問題となっている。そのサイバー犯罪を防衛するシステムに「ファイアーウォール」がある。ファイアーウォールのルールはセキュリティポリシーに沿って、手動で設定する、しかし手動設定では設定を忘れてルールは正しく作成されず、その部分について攻撃される危険がある。こうした問題に、先行研究では教師あり学習のランダムフォレストを用いることで、セキュリティポリシーに沿ったルールを自動で ACL（アクセスコントロールリスト）に適用するシステムを提案した。本研究ではさらに様々な条件のデータをランダムフォレストによって学習させ、より実用的なルールを作成するシステムを提案する。

2.3.5 潜在的なテキストのパターンを AI により健在化し、分類評価する研究（パターンランゲージの研究）

(1) 単語の局所的な部分集合を単語の連鎖と見做し、この部分集合族が得られたと仮定する。次に部分集合族の単語集合要素の評価を Boid の particle 間の強度と定義する。この様なコンセプトで人工知能の前処理 (Annotation) を設計し、Python でプログラミングして動作を確認した。

(2) Boid annotation を更に精密にする方式を先行して研究した。提案方式は Boid と LDA (Latent Dirichlet Allocation) をカスケード接続する方式である。この方式は Annotation をカスケードすること

で、Labeled-LDA の様に複雑なベイズ確率モデルを使用せず、所望の機械学習が可能になる効果を持つ。この方式を応用しセキュリティモデルに基礎付けられたテキストを教師テキストとし、評価すべきテキストを情報量として評価するシステムを提案した（研究会発表済み）。

(3)VAE(Variational Autoencoder) と GAN(Generative Adversarial Networks) 導入の計画を変更し、上記(2)に関連する LDA の研究に注力した。即ち、LDA の発展形となる Labeled-LDA, Multi-LabelLearning を調査し複数の信頼できるソースコードを入手し、動作検証、コード分析を行った。Label という構造は(2)の新提案システムとの組み合わせによる機械学習の精緻化が可能である。

(4)“ただ設計するのではなく、善く設計するとは何か？”という倫理的考察を、ウィトゲンシュタイン、ベルグソン、等を手掛かりに研究する過程に於いて、ジャン＝リュック・ナンシーの哲学が言語と身体性を連関させるという点で大きな意味を持つことが明らかになってきた。即ち、「言語ゲームとして普遍を見るとは何か」を考察する上で、分有ロゴスと身体性論は不可欠の概念であることに行き着いたのである（研究会発表済み）。

3 オントロジーを考慮した民具資料の意味的検索の研究

3.1 まえがき

近年、情報技術の発展は社会基盤を担う重要な役割を果たしている。その技術の活用は多岐の分野に渡り、生活をするうえで情報技術に触れない日はないと言っても過言ではない。情報化社会において多種多様な情報をデジタル化することは必要不可欠である。そのデジタル化の範囲は拡大しており、物体、現象、ルールなど現実世界における様々な概念のデジタル化が進んでいる。その流れの一環として、民俗学資料のデータベース構築が注目されている。これらは現物資料として保存されてきた物が多くインターネット上で管理されることにより、半永久的な情報の保存と、より簡単な資料の閲覧や検索、地域ごとの民俗学資料データとの統合を可能とする。結果的に、民俗学研究や地域産業の活性化に役立てることも可能となる。

現在福島県只見町には約 16000 点の民具が保管されており、この民具は神奈川大学 21 世紀 COE プログラムによってデータベース化され、その一部がウェブサイト「只見町インターネット・エコミュージウム」^[1]において公開されている。このデータベースを本論文では民具データベースと呼ぶことにする。民具は民俗学の立場から大いに評価される資料であるが、情報処理において扱い難い性質をもっている。^[2]したがって民具について熟知していないユーザが、民具データベースに適切な検索クエリを入力し、意図する結果を得ることは非常に困難である。また現在の民具データベースでは民具分類をはじめとする種々の情報についての意味や定義が機械可読な形式で明示されていない、例えば民具の分類方法の変更が生じた場合にデータベース上の民具を新たな基準で自動的に分類し直すということができない。そのため現在の民具データベースは民具分類の変更に対し素早く対応できる柔軟性に欠けており、結果として異なる分類方法が採用されている他の民具のデータベースとの横断検索などの相互運用が妨げられているといえる。

本研究は以上の問題を改善するために、専門知識を持たないユーザによる高度な検索や他のデータベースとの横断検索が可能な、情報の意味に基づいた処理を行う情報検索システムの構築を目的とす

る。具体的には推論による民具の自動的な分類と、それによるより高度な検索の実現を目指す。上に挙げた問題は博物館資料データベースに共通の課題であるため、最終的には本研究で提案したシステムを他の博物館の収蔵資料のデータベースにも適用することが本研究の最終目的である。したがって本研究では民具の分類方法などについては特に問題にせず、システムを実装する際に用意した民具分類はあくまで便宜上のものであることを明記しておく。

次に、同様の研究を紹介し、本研究の位置づけと新規性を明確にする。例えば文献^[3]では貴重書書誌のデータベースにおいてデータを記述する際に既存のスキーマやオントロジーで補えない部分を補完するオントロジーを構築するという試みを行っており、これによってユーザが必要とする情報をより詳細に提供することが可能となっている。この研究では本研究と同じくデータベースの情報を記述するのに Resource Description Framework (以下、RDF) を使用しており、RDF を用いたデータベースへのアプローチが意味的な検索を可能にすることを示唆している。ただし対象とする資料やその公開方法と目的が異なれば、データの形状も異なるため、これらの研究で使用された手法をそのまま本研究で用いることは難しい。そのため本研究では只見町の民具資料に適したデータの表現方法を用いる。

一方で、他の先行研究^[4]では、福島県只見町の民具データベースを対象に、民具概念の論理的な定義を行い記述したオントロジー (以下、民具オントロジー) を構築し、民具のデータ構造の提案を行った検索システムの提案がされている。この研究では概念間の関係と定義が論理的に整合性を持つことで民具を推論による自動分類を可能としている。しかし、概念定義における記述が不十分であり、論理的整合性が不完全である問題があり、推論による分類における誤った分類が生じる。また、民具を説明する農業分野などの専門的な専門用語を民具オントロジーで定義していないため、民具データベースの表現方法の制限が生じている。さらに、民具についての比較的簡単な詳細記述がなされる一般民具カードと、比較的詳しく説明がなされている国指定カードの二種類によって民具は管理されているが、国指定カードのデータベース化は実現できていない。

そこで、本研究では、生産用具に分類される国指定カードに記載される民具資料を対象とし、専門的な農作業用語を考慮した新しい民具オントロジーの構築、また民具分類における推論要素を考慮し民具データの構造を提案する。したがって、データに対してオントロジーの記述に基づいた論理演算を行い、より高度な推論機構による民具の自動分類による検索機構の構築を行う。

3.2 基礎知識

3.2.1 民具

民具とは、古くから民衆が日常生活の中で作り、使ってきた道具である。近代の大量生産された道具ではなく、人の手作業によって作られた道具を指す。民具はそこで住まう人々の日常生活と密接に関わってきた道具ゆえに、使用されていた時代の地域文化や、地形や気候の特色などを色濃く反映している^[5]。したがって、同じ道具であっても名称や形状のわずかな差が大きな意味を持つこともある。以上のことから、民具は民俗学的に重要な資料と位置づけられている。

本研究で検索データベースの対象とするのは、福島県只見町に保管される民具資料である。只見町に約 16000 点の民具資料が存在し、そのすべて、1つ1つに住民自らが、その民具についての大まかな情報を記述した「民具カード」と言われるカードがある。つまり民具カードの情報を元に、特定

の1つの民具を探ることができる。只見町の民具は大まかに「自然物採取用具」、「農耕用具」、「狩猟用具」、「漁撈用具」、「山樵用具」、「製糸用具」、「蔓細工用具」、「屋根葺き用具」の8つの分類に分けられている。「只見町インターネット・エコミュージアム」で公開されている民具データベースは、この民具カードの一部をデータベース化したものである。

この民具カードは只見町で発案された「只見町方式」と言われる記述方式に従って詳細記述がされている。この方式では、実際に道具を使っていた使用者本人が詳細記述を行う。また、備考に使用者独自の主観による自由な詳細記述をってもらう項目を設けている。従来の記述方式においては、民俗学の有識者や学者達など民具を直接使用していない人間による管理方法が主流であったが、只見町方式の整理方法では実際の民具使用者の手によって整理されている。そのため、第三者による思想の介入の余地がなく学術的に大いに評価されている整理方法である。しかしデータベース化するには非常に扱いにくい性質がある。民具カードに記載する情報は基本的には記入欄に従ったものを記載するが、その記入欄には「自由に記載して良い」とされている。また前述したように備考欄の記述はさらに自由度が高い記述がされており、記載者ごとに同じ民具資料でも表記にバラつきがあり、コンピュータが扱う情報リソースとして、そのまま使用するのは困難である。

さらに、只見町方式はマニュアル化されており、民具整理方法として採用されている他の自治体も存在する。したがって今回の只見町の民具資料を対象とした検索機構の構築は他地域の民具の検索機構構築にも応用でき、他民具データベースの横断検索にもつながるといえる。

また民具カードには一般カードと国指定カードが存在する。一般カードとは只見町の約16000点の民具全てに対して記述が行われたカードであり、国指定カードとは2333点の「会津只見の生産用具と仕事着コレクション」として国重要文化財に指定された、民具のみに新たに追加されたカードである^[6]。一般カードと、国指定カードの例を図1と図2に示す。本研究で扱うのは重要文化財の指定を受けている民具であるため、以後本論文では国指定カードのことを民具カードと呼び表す。

3.2.2 セマンティック Web

セマンティック Web とは現在の Web の延長線上に意味を明確にしたデータを用いて、コンピュータがより人間を助けられる情報空間を構築しようという試みである^[7]。構造は図3に示すレイヤーケーキで表現される。本節では、本研究で主に使用する技術、URI、RDF、RDFS、オントロジーに関するセマンティック Web の仕組みを説明する。

図1 一般民具カード

図2 国指定民具カード

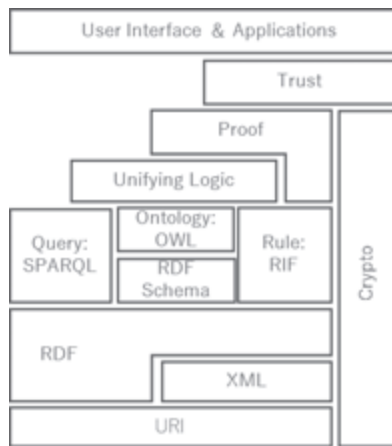


図3 セマンティック Web のレイヤー



図4 RDF トリプルのグラフ表現
Fig2.4 RDF triple Graph.

セマンティック Web では現実世界における様々な概念はリソース (resource) として扱う。リソースは web 上で識別するための名前を一意に与えられる。この名前にあたるのが Uniform Resource Identifier (以下、URI) である。したがって、URI を取得することができれば、任意のリソースを特定することが可能となる。

RDF は Web 上のリソースに関する様々なメタデータを付与するデータモデルであり、それを記述する言語体系である。セマンティック Web としてリソースで扱う現実世界における概念は多種多様である。RDF は極めて複雑な情報を記述することを可能とする。RDF における情報表現構造は主語 (subject)、述語 (predicate)、目的語 (object) の組み合わせを一単位として表される構造になっている。この構造表現を RDF トリプル (triple) と呼ぶ。図 4 に RDF トリプルのグラフを示す。リソースは主語、述語、目的語のどの部分にも割り当てることができる。RDF の文中に表現されるリソースの URI はそのグラフノードの参照と呼ばれる。URI 参照によってリソースを特定することができるため、RDF では明確な情報記述を行うことができる。またブランクノード (blank node) を用いることによって RDF トリプル自体を目的語とすることも可能である。

3.2.3 オントロジー

オントロジーとは (ontology) は本来哲学用語で「存在論」を意味する言葉である。情報科学においてオントロジーとは対象世界の概念と概念間の関係を表したもので、計算機に「知識」を与え、データの意味を理解したうえで、「知識」として処理を行う。研究分野によってオントロジーの定義や用途は様々であり、その定義は変容する。またオントロジーを記述する言語に厳密な決まりは特にない。

オントロジーは標準化されていないため、構築者の思想の違いにより、オントロジー構築の概念的な厳しさや意味論が異なる。特に記述された概念間の論理的同値関係が成立するかどうかなどの違いは、コンピュータによる推論を行う場合大きな影響をもたらす。したがってオントロジーを使用する際は、どのような分野の立場から、どのような目的でオントロジーを使用するのか明確にする必要があるといえる。また、オントロジーの定義として、「ある目的のための世界の認識に関する共通の合意」という定義が存在する。この定義からオントロジーは目的依存性が高いといえる。したがって、使用する目的と記述内容が定まれば、どの分野のオントロジーを使用し検索機構を構築するかを決定

することができる。セマンティック Web の立場におけるオントロジーは、ある情報リソースに対するメタデータの記述に際して RDF 及び RDF スキーマ（以下、RDFS）だけでは表現できない情報を記述するための語彙を定義し、提供するものである。ただし概念体系としてのオントロジー記述は、RDF 及び RDFS でも表現可能であることを明記しておく。

上述したように、どのような分野の立場、目的でオントロジーを使用するのか明確にする必要がある。本研究では、推論を用いた検索機構のシステム構想と、民具カードの詳細は自由な記述で表現されている 2 つの点から、セマンティック Web の立場からオントロジーを使用している。その根拠については、次章にて詳しく言及する。またセマンティック Web では URI の導入や記述論理 (Description Logic) をベースにしたオントロジー言語も用い、概念の具体的な明記と、概念間の論理関係の記述を実現し、オントロジー記述に基づいた推論機構の構築に適している。

また、セマンティック Web の分野のみならず、すべてのオントロジーの分野に共通して、概念化に関する暗黙的な情報を明示し、対象世界における概念の定義や概念間の関係を機械可読な形で記述することで、知識の共有・再利用や意味に基づいた情報処理を可能にするという利点があるといえる。

3.2.4 Agriculture activity ontology

AGROVOC とは、国際連合食糧農業機関と欧州共同体委員会が共同で開発した農林水産、食糧、環境、およびその関連分野の専門用語 32,000 語以上を網羅したシソーラスであるが、日本の農作業名称における標準語彙体系として既存の語彙体系は概念間の関係の曖昧性、農作業名称が持つ語彙の多様性を表現できなかった。これらの課題解決のため農作業の概念間の関係をより細分化し、また概念と用語表記を分離することで、明確に概念を定義し、概念間の関係を構造化したオントロジーが Agriculture activity ontology(以下、AAO)^[10]である。

先行研究における民具オントロジーは概念系を独自に考察し構築したものである。しかし、対象とするデータの多様性に応じオントロジーも拡大するため、生じる概念関係が複雑化する。そこで、本研究では、既存のオントロジーである AAO を採用する。AAO を利用することで、オントロジーの論理的整合性を維持し、専門的な農作業用語を機械が可読できるメリットがある。しかし、AAO では近代農業における農作業も定義しており、民具資料検索を行う上で不必要な概念定義も存在する。よって、本研究では、AAO が定義する農作業名の上下関係と、農作業語彙の定義を一部抜粋し本研究で構築するオントロジーに組み込む。

3.2.5 RDF/XML

RDF とは Web 上にあるリソースのメタデータを記述するための枠組みである。実際に RDF で表現されたデータをコンピュータに認識させるためには、RDF のグラフをテキストの形式に直列化 (serialize) する構文が必要になる。RDF/XML はこうした構文の 1 つであり、RDF 用の XML 構文である。RDF を直列化する構文には他にも Notation3 や Turtle などの複数の種類があるが、RDF/XML には既に広範に普及している XML 用のツールが使用できるという利点や XML ベースのシステムと相互運用が比較的簡単であるという利点が存在する。そのため本研究では RDF の記述にこの RDF/XML を使用する。

3.2.6 OWL

Web のリソースを記述するオントロジー言語として 2004 年に Web 関連技術の標準化を行っている W3C(World Wide Web Consortium) によって勧告されたのが OWL (Web Ontology Language) ^[7] である。OWL には、論理記述に基いて厳密に OWL Lite、OWL DL、OWL Full の 3 つのサブ言語に区別される。OWL Lite は 3 つのサブ言語のうち最も使用できる語彙が制限されたシンプルな言語である。したがってクラス階層や基本的な制約条件の記述しか行うことができないが、記述論理に基づいた論理計算の完全性と決定可能性は保証されるという特徴がある。

OWL DL は OWL の語彙を全て利用することが可能であり、なおかつ論理計算の完全性と計算可能性が確保されるように設計されている。したがって厳密な推論の実行とある程度の表現力を両立しているサブ言語であるといえる。OWL Full は最も表現力が高いサブ言語であるが、自由な記述を行える反面、OWL DL などと異なり論理計算の完全性と計算可能性が保証されないという特徴がある。よって OWL Full によって記述論理に基づいた推論を行うことができるオントロジーを構築するのは難しいといえる。本研究では記述論理に基づいた論理計算の完全性と決定可能性が保証されるという点と語彙が制限されないという点から、OWL DL のレベルでオントロジーの記述を行う。

3.2.7 SPARQL

SPARQL とは RDF の形式で表現されたデータの検索を行うための問い合わせ言語 (query language) の 1 つである。RDF のための問い合わせ言語には他にも N3QL などが存在するが、SPARQL は 2008 年に W3C 勧告となっており、今後使用機会のより一層の増加が見込まれる言語である。

3.2.8 Protege

Protege はスタンフォード大学で開発されているオントロジーエディタであり、OWL によるオントロジーの記述を完全にサポートしている。Protege は豊富なプラグインによってユーザによる機能拡張が容易に行えるなどの利点から、オントロジーエディタとして世界的に普及している。現在 Protege には通常のエディタである ProtegeDesktop と複数人でのオントロジーの共有が可能な WebProtege の二種類があるが、本研究ではオントロジーの構築に ProtegeDesktop を用いる。

3.2.9 Jena

Jena(Apache Jena) はセマンティック Web や Linked Data を扱うアプリケーションを Java によって実現するためのフレームワークである。Jena が提供する機能として RDF によって記述されたデータの入出力や、OWL によって記述されたオントロジーに基づいた推論、SPARQL による検索などがある。本研究におけるシステムの実装には Jena を用いる。

3.3 提案システム

3.3.1 セマンティック Web を用いたオントロジー

前述したとおり、オントロジーの定義や記述言語は標準化されていない。よって本研究におけるオ

ントロジーの使用位置づけとオントロジー記述言語に OWL を採用した理由を説明する。そもそも本研究においてオントロジーを使用する目的は、オントロジーの記述に基づいた推論機構を用いることで、RDF 化された民具データのメタデータを意味的に処理し、民具を自動分類することが目的である。よって本研究におけるオントロジーの主な記述内容は、民具の分類定義と民具の持つメタデータの定義ということになる。つまり、本研究に使用するオントロジーには、推論機構を実現できる技術の確立と、民具の情報を表現するためのある程度自由な記述が許容されることが求められる。推論を用いた検索機構のシステム構想と、民具データを自由に記述で表現という2つの観点から、本研究におけるオントロジーはセマンティック Web の立場におけるオントロジーを使用する。

3.3.2 提案システム

本研究が提案する情報検索システムは、RDF 化した民具データベースと、データベース上の情報について、意味と定義を農作業の概念を考慮して記述した民具 AAO の2つの部分から構成される。このシステムにユーザが SPARQL による問い合わせを行うと、RDF で表現された民具の情報に対して、農作業名や民具分類の定義などが記述された民具 AAO に基づいた推論が行われ、推論によって導かれた論理関係を踏まえた検索結果が返される。この仕組みを図にしたものを図5に示す。

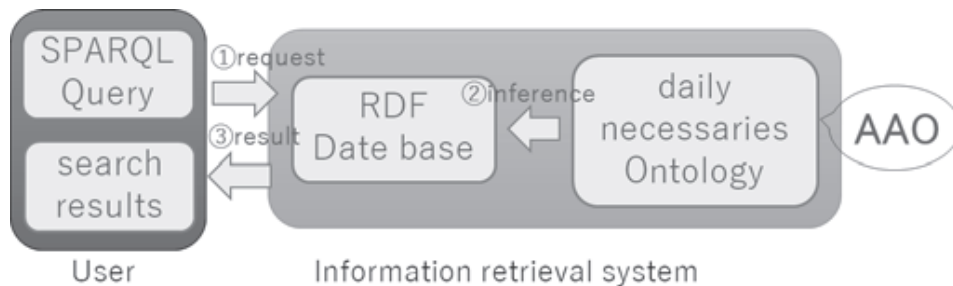


図5 提案システム

只見町での民具の分類方法は、民具の使用用途に応じた分類方法を採用している。分類の定義を行うためには民具カードの項目のうち、使用方法の項目（以下、詳細）に該当する情報が必要である。したがって本研究では最低限必要となる民具カードに記載された民具の名称、分類番号、詳細を RDF で記述することにした。ただし使用目的として民具カードに記載されている文章はそのまま文字列として表現しても民具分類を論理的に定義することができないため、詳細に記述されていた説明文からその民具の、農作業名、使用場所、使用時期、使用対象物、使用方法の5つに細分化し、これら5つの要素をブランクノードを用いて、まとめて「詳細」というプロパティのオブジェクトとして表現することにした。これは、複数の用途を持つ民具であっても、その用途の数と等しい数のブランクノードを「詳細」のオブジェクトとして使用し同様の記述を行えば、用途ごとの使用対象や役割の混乱なく民具の情報を表現することができるという利点が存在する。また提案システムにおいては、リソースである民具の URI を決定するのに、民具と一対一である分類番号を利用している。こうすることで同一名称の民具であっても明確に区別することができる。このとき民具の名称は、リソースとしての民具をサブジェクトとし、「名称」というプロパティのオブジェクトとしてリテラル（文字列）

で表現される。民具の分類番号も同様に「分類番号」というプロパティのオブジェクトとしてリテラルで表現される。この民具のデータを RDF グラフで表現したものを図 6 に示す。



図 6 提案システムにおける RDF グラフ表現

3.4 提案システムの実装

提案した情報検索システムの実装は Jena を用いて Java 言語で行った。使用したソフトウェアの一覧を以下に示す。Jena は Eclipse 上で使用している。

- Protege 5.2.0 Beta
- Eclipse Kepler Service Release
- Jena Version 2.7.4

3.4.1 民具データベースの RDF 化

まず使用した民具のデータについて説明する。本研究で使用した民具のデータは只見町で保管されている民具うち、耕土制御を目的とする民具 8 点と、実験用に作成した架空の民具である「テスト」1 点の、合計 9 点を使用する。これらは RDF に記載される詳細の内容を推論し分類を行うため、民具カードに記載される情報のうち名称（地方名もしくは別名）、分類番号、詳細のみに限定して RDF で記述を行う。また民具を表す URI の一部に民具に割り振られる分類番号を利用することで、URI を一貫性を保持し、URI の重複防止を保証している。更に詳細記述に基づいた分類を行うために、民具カード上の詳細の説明文の他に、その民具の農作業名、場所、時期、対象物、方法に対する役割を名詞と動詞の形で抜き出した。実際の民具の分類に使用したのは詳細のうち農作業名、場所、方法の 3 項目である。使用した民具のデータの一覧を表 1 に示す。本研究では、これを実験を行う上での仮の民具データベースとして扱う。

表 1 RDF 化した民具

分類番号	名称	詳細	農作業	使用場所	使用方法
A4430	カイリョウグワ	田の耕起やシロカキを行う道具	耕起, 代かき	田	人
A354	シロカキマンガ	水田の耕起を行う道具	耕起	水田	牛, 馬
A396	タゴシラエクワ	湿田を耕す道具	耕す	湿田	人
A1169	バコウグワ	乾田を耕す道具	耕す	乾田	牛, 馬
A3104	サイドキ	田の代かきを行う	代かき	田	牛, 馬
A1262	ヒラグワ	田の耕起を行う道具	耕起	田	人
A2918	クワガラ	田を耕す道具	耕す	田	人
A34	ニダンコウ	田の耕起を行う道具	耕起	田	牛, 馬
M	テスト	乾田のテジロカキを行う道具	テジロカキ	乾田	人

次に民具 RDF のメタデータの記述構造について説明する。民具 RDF の記述は基本的に図 6 になる。

実装における記述では、プロパティは全て英語表記となっている。具体的には、名称を「name」、分類番号を「id」、詳細を「description」、農作業を「AA」、場所を「field」、時期を「season」、方法を「method」としている。またリテラルとして記述されている名称を除く、プロパティは AAO が定義する英字表記で記述し、AAO が定義していない「テジロカキ」などといった農作業名などはローマ字で記述されている。例として第 1 分類番号が A1262 のヒラグワという民具の RDF グラフによる表現を、図 7 に示す。ここで図中における [ex:] は本研究における名前空間「http://example.com/test#」を表す名前空間接頭辞である。例えば、図 7 においてヒラグワ自身を表す URI 「ex:A1262」は「http://example.com/test#A1262」を意味している。更に RDF 化したヒラグワの RDF/XML による一部記述を図 8 に示す。



図 7 民具 RDF グラフ表現の例

```
<rdf:Description rdf:nodeID="A2">
  <ex:text rdf:datatype="http://www.w3.org/2001/XMLSchema#string">田や苗代などのテジロカキ</ex:text>
  <ex:act rdf:resource="http://example.com/test#Puddling1"/>
  <ex:method rdf:resource="http://example.com/test#human"/>
  <ex:field rdf:resource="http://example.com/test#Rice_nursery"/>
  <ex:field rdf:resource="http://example.com/test#Paddy_field"/>
</rdf:Description>
<rdf:Description rdf:about="http://example.com/mi#M1262">
  <ex:description rdf:nodeID="A2"/>
  <ex:id rdf:datatype="http://www.w3.org/2001/XMLSchema#string">A1262</ex:id>
  <ex:field rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ヒラグワ</ex:field>
```

図 8 ヒラグワの RDF/XML による記述

3.4.2 AAO を考慮した民具オントロジー

本研究では AAO の農作業概念の語彙体系を一部引用し、民具オントロジーを再構築する。AAO には近代農作業における手法など様々な農作業名の語彙が定義されている。しかし、本研究で扱う民具資料は近代農作業における概念要素を持たず、推論による民具分類には不必要である。よって民具資料の分類において必要な概念系とその定義のみを AAO から参照するものとする。分類方法は文化庁分類と同じく民具の使用目的による分類とする。本研究において民具の詳細と民具の使用目的は同義である。構築する民具オントロジーには分類の定義と、その定義を論理的に行うための概念の記述を行う必要がある。本研究では実験場の仮の分類として「鋤」という分類を定義した。またそれらの上位概念として「民具」という概念を定義した。それぞれの定義を以下に示す。

- ・鋤： 使用場所が「田」もしくは「畑」、かつ使用動作が「耕す」、かつ使用方法が「人」であるモノ
- ・民具： 何らかの役割を持つモノ

続いて Protege によって構築した AAO を考慮した民具オントロジーの構造を図 9 に示す。ただしこの図は分かり易さのためにクラスやプロパティをラベル (label) によって日本語で表記したものであり、この民具オントロジーの構成要素の本来の名前、すなわち URI は英語またはローマ字で記述されている。名前空間は RDF 化した民具データベースと同じく「http://example.com/test#」である。また民具オントロジーの構成要素の一覧を表 2 に示す。

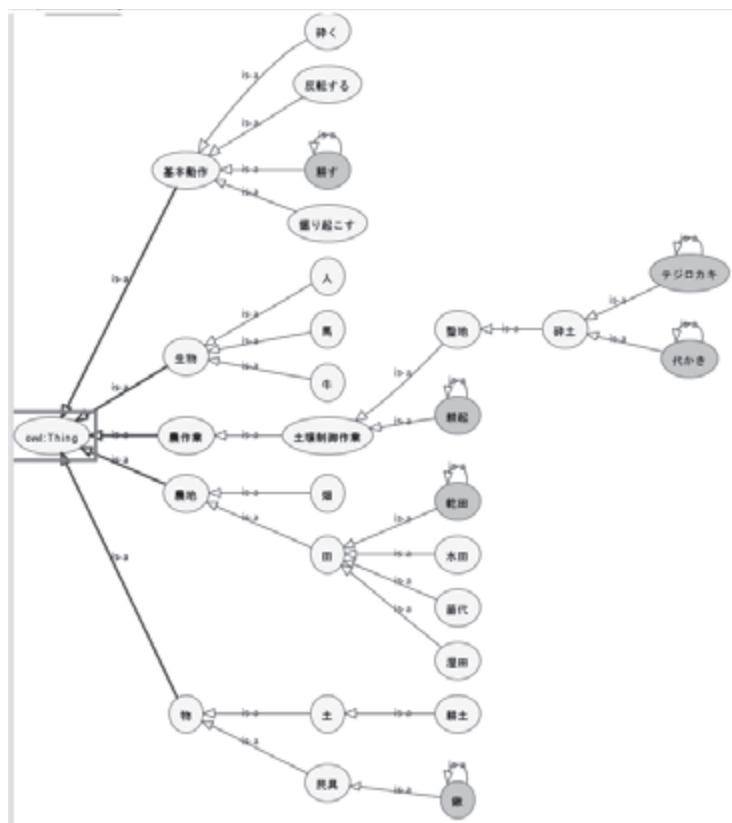


図9 構築した民具 AAO

表2 構築した民具 AAO の構成要素

ラベルによる 日本語表記	実際の名前	種類
モノ	mono	クラス
民具	daily_necessaries	
鋤	hoe	
田	Paddy_field	
湿田	Wet_paddy_field	
乾田	well_drained_paddy_field	
苗代	Rice_nursery	
耕起	Plowing	
代かき	Puddling	
テジロカキ	Tejirokaki	
砕土	Pulverization	
人間	human	
牛	cow	
馬	horse	
詳細	description	プロパティ
場所	field	
農作業	AA	
方法	method	
対象物	object	
時期	season	

次に Protege にて定義した「鋤」について図 10 にて示す。これは「田」もしくは「畑」というインスタンスを「filed」というプロパティのオブジェクトとして、どちらか片方、もしくは両方を持ち、「耕す」というインスタンスを「AA」というプロパティのオブジェクトとして持ち、「人」というインスタンスを「method」というプロパティのオブジェクトとしてもつようなインスタンスの集合は「鋤」に定義されることを記述している。

Description: 鋤
Equivalent To +
● ((filed some 田) or (filed some 畑)) and (AA some 耕す) and (method some 人)

図10 protege における鋤の定義

Description: 耕起
Equivalent To +
● ((act some 反転する) or (act some 掘り起こす)) and (object some 耕土)

図11 protege における耕起の定義

次に AAO の概念定義を参照し、「耕起」、「テジロカキ」、「代かき」の農作業名にあたる 3 つのクラスの定義と、「耕す」の基本動作にあたるクラス 1 つを定義した。そのそれぞれの定義を以下に示す。

- ・耕す：使用対象が「耕土」かつ使用動作に「掘り起こす」または「反転する」または「砕く」のいずれか 1 つ以上を持つ
- ・耕起：使用対象が「耕土」かつ使用動作に「掘り起こす」または「反転する」のいずれか 1 つ以上を持つ
- ・代かき：使用対象が「土」かつ使用動作に「砕く」を持つ

- ・テジロカキ: 使用対象が「土」かつ使用動作に「砕く」を持つ

また、例として Protege にて定義した「耕起」について図 11 にて示す。

3.5 実験及び実験結果

3.5.1 民具 AAO における推論の実験及び実験結果

提案する情報検索システム全体に対する実験を行う前に、4.2 で提案した民具 AAO における民具分類の定義が推論を行った時に適切に処理を行うか確認する必要がある。そこでまず図 9 で示す民具オントロジーに、適当な架空の民具を「民具」クラスのサブクラスとして記述し、推論を行った時に記述された論理関係に従って民具が分類されるかを確認するという実験を 2 種類に分けて行った。推論は protege のプラグインである推論機構 HermiT を使用する。まず、民具オントロジーの分類が適切に処理されているかを確認するため、分類定義を追加した架空の民具の一覧を表 3、4 に、推論を行う前の民具オントロジーの構造を図 12 に示す。また実験結果、つまり推論を行った後の民具オントロジーの構造を図 13 に示す。

表 3 民具 AAO に追加した架空の民具

名称	農作業	使用場所	使用方法
鋤 1	耕す	田	人
鋤 2	耕す	畑	人
鋤 3	耕す	田かつ畑	人

表 4 民具 AAO に追加した架空の民具

名称	農作業	使用場所	使用方法
鋤 4	耕起	田	人
鋤 5	耕す	乾田	人
鋤 6	耕す	田	牛

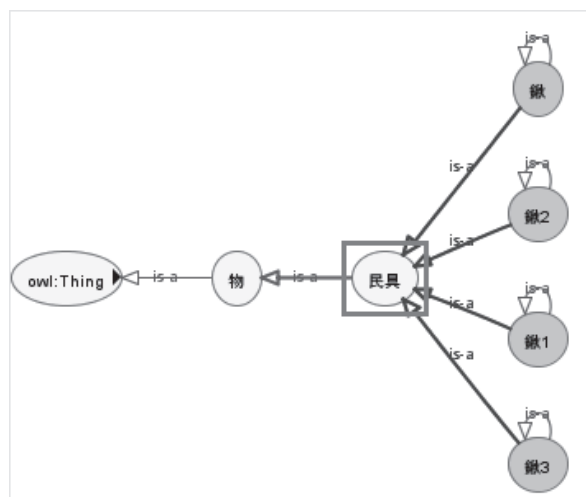


図 12 推論前の民具分類

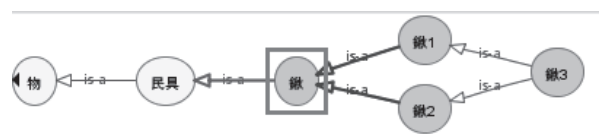


図 13 推論後の民具分類

図 12 と図 13 の民具 AAO 構造を比較して推論処理の説明をする。オントロジー内の記述に基づく推論を行った結果、鋤 1、鋤 2 は「鋤」のサブクラスとなり、鋤 3 は「鋤 1」と「鋤 2」の両方のサブクラスになっている。図 10 にて示したように、「鋤」とは「田」と「畑」のどちらか片方、もしくは両方を「耕す」、「人」が使う民具と定義している。表 3 で示した、架空の民具である鋤 1、鋤 2

はそれぞれ、「田」を「耕す」、「人」が使う民具と「畑を」を「耕す」、「人」が使う民具である。よって推論後に「鍬」のサブクラスとして記述された「鍬 1」と「鍬 2」は分類定義にしたがって適切に分類されている。次に鍬 3 では「田」と「畑」両方を「耕す」、「人」が使う道具であると定義されている。よって使用場所が「田」である鍬 1 と使用場所が「畑」である鍬 2 のサブクラスと推論されている。また前述したとおり、鍬 1 と鍬 2 は「鍬」のサブクラスであり、「鍬」は鍬 1、鍬 2 のサブクラスを間接的にサブクラスとみなすことができる。よって鍬 3 においても適切な分類が行われているといえる。

次に、分類定義と異なるインスタンスを持つ民具に対し、民具 AAO がインスタンスの推論から適切に民具分類を行えるのかを確認する。分類定義を追加した架空の民具の一覧を表 4 に、推論を行う前の民具オントロジーの構造を図 14 に示す。また実験結果、つまり推論を行った後の民具の分類構造を図 15 に示す。また、同時に農作業の推論前と推論後の概念構造を図 16 と 17 に示す。

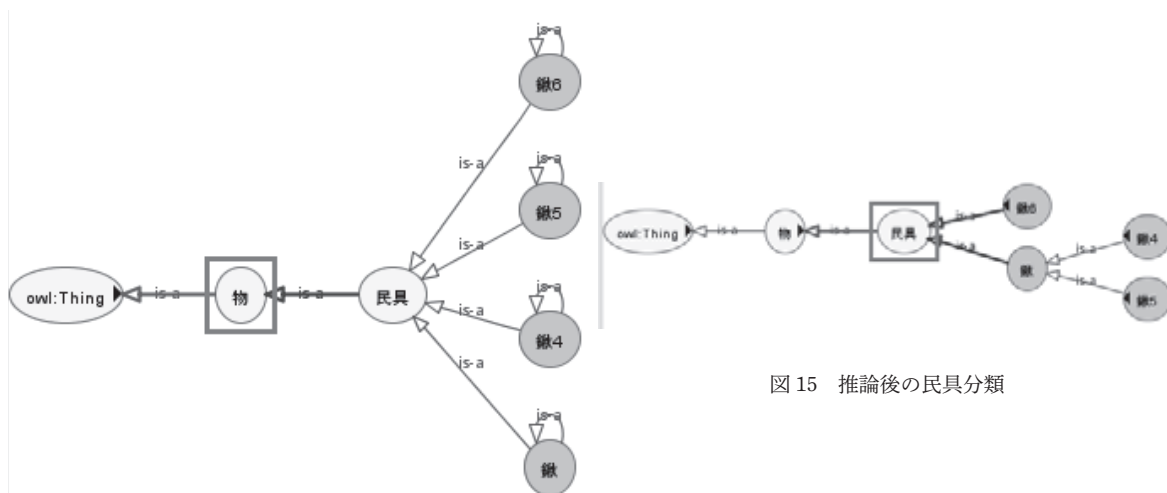


図 14 推論前の民具分類

図 15 推論後の民具分類

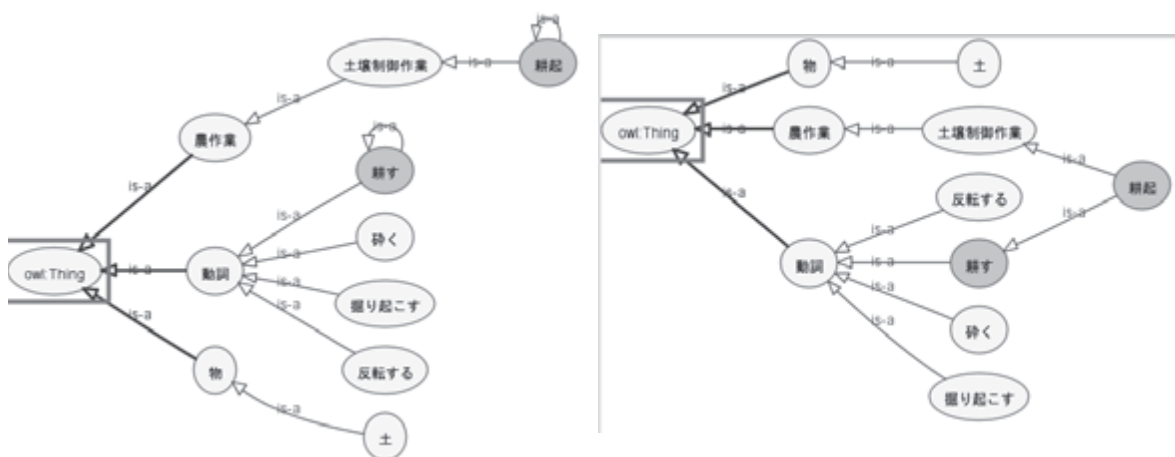


図 16 推論前の農作業の概念構造

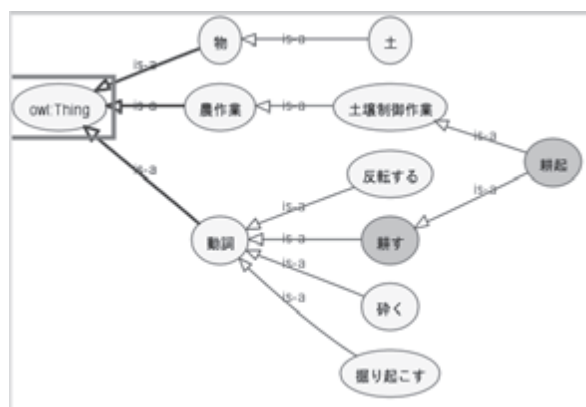


図 17 推論後の農作業の概念構造

図 14 と図 15 の民具 AAO 構造を比較して推論処理の説明をする。「鍬」定義は前述と同様である。「鍬」のサブクラスは鍬 4、鍬 5 になっている。鍬 6 に関しては民具のサブクラスとして推論前と推

論後で変化は生じていない。まず鍬 4 は、「鍬」の分類定義とは農作業のインスタンスの名称が異なっているのかかわらず、鍬に分類されている。ここで図 16 と図 17 を参照すると推論前と推論後で「耕起」が「耕す」のサブクラスに分類されているのが分かる。前章で示したように、「耕起」は使用対象が「耕土」かつ使用動作に「掘り起こす」または「反転する」のいずれか 1 つ以上を持っていると定義している。一方、「耕す」は使用対象が「耕土」かつ使用動作に「掘り起こす」または「反転する」または「砕く」のいずれか 1 つ以上を持つと定義しており、「耕起」というクラスは「耕す」というクラスであるための条件を満たしているため推論による分類が行われた。よって、鍬 4 は「田」を「耕す」、「人」が使う民具という定義に置き換えることができる。次に、鍬 5 について使用場所が「乾田」になっているものの「鍬」に分類されている。民具 AAO 内では、「田」のサブクラスとして乾田、水田、湿田などが記述されている。したがって民具 AAO の概念世界では、「田」というインスタンスは乾田、水田、湿田といったインスタンスの総称であると言い換えることができる。よって鍬 5 が「鍬」に分類されることは適切であるといえる。鍬 6 は使用方法のインスタンスが牛になっており、「鍬」に分類されておらず適切な処理が行えているといえる。

3.5.2 提案システムに対する検索の実験及び実験結果

実装した提案システムにおいて、検索結果にオントロジーの記述が反映しているか、確認を行う実験を実施した。実験では表 1 に示す、意味的に「畑や田を耕す」民具に該当する民具に加え、テスト用に作成した架空の民具の計 9 点の民具を RDF で記述した民具データベースに対して、既存研究 [1] で提案された民具オントロジーを導入した場合と、図 9 に示す、本研究で作成した民具 AAO を導入した場合の二通りを、それぞれ SPARQL による検索クエリで問い合わせる手法をとる。SPARQL によって記述された、「鍬」の民具を問い合わせる検索クエリを図 18 に示す。

```

PREFIX ex: <http://example.com/test#>
PREFIX mingu: <http://example.com/mi#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?min ?name ?id
WHERE{
  ?min ex:name ?name. ?min ex:id ?id. ?min ex:description ?blank. ?blank rdf:type ex:鍬}

```

図 18 SPARQL における「鍬」に分類される民具を検索するクエリ

実験結果は、「鍬」という分類名に対して、既存手法の民具オントロジーを導入した場合の検索結果を図 19 に示す。次に、「鍬」という分類名に対して、本研究の民具 AAO を導入した場合の検索結果を図 20 に示す。

min	name	
mingu:A2918-2-1-168	"カワガラ"^^<http://www.w3.org/2001/XMLSchema#string>	
mingu:A396-2-1-133	"タゴシエカワ"^^<http://www.w3.org/2001/XMLSchema#string>	
mingu:A1169-2-1-208	"バコウガラ"^^<http://www.w3.org/2001/XMLSchema#string>	

図 19 民具オントロジー参照した場合の検索結果

min	name	i
mingu:A2918-2-1-168	"クワガラ"^^<http://www.w3.org/2001/XMLSchema#string>	
mingu:M1262-2-1-179	"ヒラグワ"^^<http://www.w3.org/2001/XMLSchema#string>	
mingu:M	"テスト"^^<http://www.w3.org/2001/XMLSchema#string>	
mingu:A4430-2-2-9	"カイリョウグワ"^^<http://www.w3.org/2001/XMLSchema#string>	
mingu:A396-2-1-133	"タゴシラエクワ"^^<http://www.w3.org/2001/XMLSchema#string>	

図 20 民具 AAO を参照した場合の検索結果

実験結果を比較する。まず「鋤」に分類される民具を検索した結果である図 19 と図 20 を比較する。既存研究のオントロジーを導入した際の実験結果では図 19 のように「クワガラ」、「タゴシラエクワ」、「バコウグワ」という民具が検索結果として出力された。出力された民具はいずれも「耕す」というクラスを持っており、その他の民具は出力されなかった。これは、民具オントロジーでは「耕す」というクラスの意味の推論が実現できていないため、意味的に「耕す」というクラスに包括されるクラスが付与されている民具を推論できていないことが分かる。しかし、「タゴシラエクワ」や「バコウグワ」のように「水田」、「湿田」といったクラスが付与されている民具が出力されていることから、「田」というインスタンスは乾田、水田、湿田といったインスタンスの総称であるとシステム内で認識していることが分かる。

次に、民具 AAO を導入した際の実験結果では図 20 のように「クワガラ」、「ヒラグワ」、「テスト」、「カイリョウグワ」、「タゴシラエクワ」の 5 つの民具出力された。新たに出力された「テスト」、「カイリョウグワ」、「ヒラグワ」はいずれも「耕す」とは異なるクラスが付与されているが、民具 AAO 内で意味的に「耕す」というクラスに包括されるクラスが付与されている民具を推論しており、使用方法を参照することで「バコウグワ」などの牛、馬などが使用する民具などは検索結果から除外されており、適切な分類結果が返ってきたといえる。

よって以上の実験結果から、本研究で提案したシステムは、民具データベースに対して民具 AAO の概念系に基づいた推論による適切な自動分類が実装されたと評価できる。

3.5.3 考察

3.5.4 提案システムの評価

まず 3.5.1 のオントロジーにおける推論の実験結果から、築した民具 AAO では推論により分類の定義に従って民具を自動的に分類することができることが確かめられた。また、民具 AAO の農作業語彙の定義を参照し、意味的等価であると結論づけられる農作業名を紐づけること、分類定義を推論することで任意の民具データを推論する二重推論を可能にしている。よって、生産用具や民具に特化したより高度な検索機構の構築がなされている。加えてこの民具 AAO を導入した場合と、民具オントロジーを導入した場合で提案システムに対する各分類の民具の検索結果を比較した 5.2 の実験では、民具 AAO を採用したシステムでのみ農作業概念を参照した、二重推論による広範囲推論と、論理的に整合性をもつ適切な分類である民具の検索結果が得られた。

よって、本研究で提案した農作業概念を考慮した民具 AAO を参照した情報検索システムは、民具の分類定義に該当する任意の民具をメタデータが一致しているかで識別するのではなく、メタデータの論理的整合性による識別で検索システムを構築しているといえ、従来の民具データベースに比べて

より高度な検索を行うことが可能であると評価できる。また本研究では扱わなかったが、狩猟用具、漁撈用具、三樵用具、製糸用具、蔓細工用具、屋根葺き用具に分類される只見町の民具資料に関しても、各分野に特化したオントロジーを本研究と同様に組み込むことで、民具データベース全体の推論機構の高度化を図ることができると予想される。

3.6 結論

本稿では只見町の民具データベースを対象に専門知識を持たないユーザが任意の検索結果を得られることを目的とし、AAO に基づいた農作業概念を民具オントロジーに組み込み、民具 DB を RDF 化し、意味に基づいた情報処理を行う、意味的検索システムの提案と実験を行った。実験結果から、提案システムでは、専門的な農作業用語で説明される、または只見町の方言で説明記述される民具に対して、民具 AAO を参照し、推論を行うことで、民具の適切な自動分類を行い、検索を行うことが可能であることが確かめられた。以上の結果から、本研究の目的はおおむね達成されたと結論づける。

しかし、本研究では農耕用具の民具に着目した検索機構である。一方で、只見町には狩猟用具、漁撈用具、三樵用具、製糸用具、蔓細工用具、屋根葺き用具、の分類される民具が存在する。したがって、全民具を対象にした検索システムとしては不十分であるという問題があり、実用化を考える上で今後、解決すべき課題である。

また、これまでに述べてきたように、民具などの博物館資料はその種類、分野を問わず、学術的にも文化的にも重要なものである。しかし、高齢化に伴い博物館資料を実際に使用していた有識者は年々減少し、それと同時に博物館資料がもつ情報を読み取ることが不可能になる危険性がある。民具の情報を後世に半永久的に残すため、デジタルミュージアムのような取り組みは、今後増加すると予想される。本研究はその取り組みの一助となることを目指して博物館資料に対する情報検索システムを提案するものであったが、例えば、福島県只見町では地域経済の活性化の一環として民具資料の活用を検討しており、情報の検索だけでなくその情報の提供の仕方も考慮する必要がある。将来的にはこうしたシステムの利用目的を強く意識した立場からの研究を行うことが、そのシステムの実用性を高める上で重要になってゆくと考えられる。

4 情報カプセルとブロックチェーンを用いたデジタルコンテンツの流通

4.1 まえがき

近年、ネットワークの高速化やスマートフォンの普及により、音楽、動画、画像などのデジタルコンテンツが手軽なものとなり、流通が盛んになってきている。そこでデジタルコンテンツの普及に伴い様々な DRM（デジタル著作権管理）が提案されているが、従来の DRM だとコンテンツ配信業者がコンテンツ利用を制限して、利用者同士でのコンテンツの譲渡を制限されてしまっている。また、ブロックチェーンの登場によりブロックチェーンでの著作権管理も提案されている。しかし、著作権だけではなく個人情報に関する権利の保護が必要であり、情報セキュリティの対策も不可欠なものになっている。これらを改善するには、利用者同士でのコンテンツの譲渡を可能にし、著作権や個人情報に関する権利を保護する情報セキュリティ対策を行ったコンテンツ流通システムが求められる。^{[11][27]} デ

デジタル化されたコンテンツがオリジナルか二次利用されたコンテンツであるかという証明、不正使用対策には従来の DRM での実現は困難である。先行研究では、信頼できる第三者により解決しているが、プライバシー、結託、第三者の存続性の問題があるため完全に信頼することが難しい。また、先行研究では利用者同士でのコンテンツ譲渡をするためにスマートプロパティと既存の DRM 技術を組み合わせることによるコンテンツのコピー防止の実現が提案されているが、システムのセキュリティモデルが示されていない^{[12][13][14][18]}。デジタルコンテンツの流通に暗号化されたデジタルコンテンツと、それに対する暗号鍵、著作権情報、管理エージェントなどをパッケージ化した情報カプセルを導入し、信頼できる第三者を必要としない方法を提案するために、仮想通貨に用いられているブロックチェーン技術を用いる。これらを組み合わせることで、従来の情報カプセルのエージェント間で通信を行う際に流通後のコンテンツの位置の把握が困難ということや、利用者などのプライバシーを保護する必要があるという問題点を解決できると考えられる。また、流通システムのセキュリティモデルに Take-Grant^{[16][17]}モデルを採用し、よりセキュアなシステムの実装を目指す^{[15][23]}。本研究では、ブロックチェーンを情報カプセルとみなし、ブロックチェーン技術で可能となるスマートコントラクトに Take-Grant モデルを記述し、デジタルコンテンツの譲渡などのやり取りにおいて利用権や契約情報が改ざんされていないかの健全性を示すことを目的とする。4.2では基礎知識、4.3では提案手法、4.4では実装した結果、5章では本研究の結論を示す。

4.2 準備

4.2.1 情報カプセル

情報カプセルは、デジタルコンテンツ自体と、コンテンツの表現に関するもの、コンテンツの利用を制御・制限するもの、コンテンツの利用制御を行う機能の制御をするものなどをカプセル化したものである。コンテンツの制作者によって利用条件が設定され、利用回数や期間、用途など様々な項目でユーザの利用を制御・制限することが出来る^{[12][25]}。

4.2.2 Take-Grant モデル

Take-Grant^{[16][17]}モデルとは、最も一般的なセキュリティモデルであるアクセス制御行列モデルである。特別なポリシーには依存せず保護状態を端的に表現することができる。HRU(Harrison, Ruzzo, Ullman:1976)^{[28][29]}モデルはアクセス制御行列を用いてポリシーの仕様からその安全性を証明する一般的な方法はないことを示すのに使われた。これに対し、Take-Grant モデルは安全性が証明できるモデルとして提案された。Take-Grant モデルは、特定のシステムを理論的および実用的側面から分析するための方法を与える。このモデルは、権限の委譲に関わる操作 take と grant を含む基本操作を設定し、その基本操作を持つシステムのセキュリティをグラフ理論に基いて分析する。このような枠組みの中では、ある主体が他の主体（あるいは対象）に対して特定の権限を持つかどうかを決定するアルゴリズムが存在することが示された。Take-Grant モデルではシステムの保護状態は、主体（および対象）を頂点（vertex）、許可をラベル付きの辺（edge）とする有限の有向グラフ（directed graph）によって表される。また、システムの基本操作はグラフに作用する書換え規則によってモデル化され、保護状態の変化はグラフの書換えに対応付けられる。システムの安全性判定は、初期状態に基本操作を適

用して得られる保護状態において、主体 p が主体（あるいは対象） q に対して権限 α を持つかどうかを判定することとして形式化される。[16] で示された take-grant システムは、次の 5 個の基本操作を持つ。

- (a) Take: 一定の条件のもとで、主体が他の主体（または対象）の持つ権限を獲得する。
- (b) Grant: 一定の条件のもとで、主体が他の主体（または対象）に権限を委譲する。
- (c) Create: 新たな主体を作成する。
- (d) Call: 主体がプログラムを呼び出すとき、プロセスにあたる主体を作成する。
- (e) Remove: 主体が持っていた権限を削除する。

主体のみから構成されるシステムでは、主体 p が主体 q に対して権限 α を持つことを、グラフの大きさに比例する時間で判定するアルゴリズムが存在することが示された。しかし、主体のみから構成されるシステムは制限が弱く、ユーザ p が q を read（あるいは write、call）できるならば、 p あるいは q と直接／間接に何らかの許可関係を持つすべてのユーザは、基本操作を繰り返し適用することで q を read（それぞれ write、read と call）する権限を得ることができる。^{[16][17][24]}

現実のシステムの多くは主体と対象を区別しており、このようなシステムは S-O take-grant システムによってモデル化される。ここで、主体の集合と対象の集合は互いに素である。S-O take-grant システムでは、 p が q に対して権限 α を持つことを判定するアルゴリズムが存在することが示された。

[16] では、表記上の不整合はあるが、モデルは数学的な厳密さのもとに規定されている。しかし、その記述は形式言語によって行われたものではない。

4.2.2.1 主体のみから構成されるモデル

システムの保護状態は、次の要素からなる有限の有向グラフ（以下では、保護グラフと呼ぶ）によってモデル化される。

- ・頂点: 主体
- ・ラベル付きの有向辺: 主体が持つ許可。ラベルは r 、 w 、 c を要素とする空でない集合であり、権限を表す。

ラベルは形式的なオブジェクトであるが、インフォーマルには r は read、 w は write、 c は call と解釈できる。すなわち、頂点 x から y への辺が存在し、 r がラベルに含まれるならば、 x は y を read できることを意味する。

システムの基本操作は、有向グラフに対する書き換え規則としてモデル化される。

システムの安全性分析問題は、保護状態を表す有向グラフ G と保護違反を表す特性 X 、基本操作を表す書き換え規則の集合 R が与えられたとき、 $G = G_1$ かつ G_n が特性 X を持つようなグラフの列

$$G_1, G_2, \dots, G_n$$

が存在するかどうかを判定する問題として形式化される。ここで、 G_{i+1} は R 中の書き換え規則を G_i に適用することによって得られるグラフ。特性 X は、次のように表される。

グラフの頂点 p から q へ、ラベル α を持つ辺が存在する

以下ではこれを、 p が q に対して α 可能であると言う。

take-grant システムの基本操作に対応する書き換え規則は、次の 5 個である。

(a) Take

グラフ G に相異なる頂点 x, y, z が存在して条件

- x から y へのラベル γ の辺が存在し、 $r \in \gamma$ である
- y から z へラベル $\alpha \subseteq r, w, c$ の辺が存在する

を満たすとき、 G を次のように書き換えたグラフ G' を生成する。

- x から z へラベル α の辺を追加する

(b) Grant

グラフ G に相異なる頂点 x, y, z が存在して条件

- x から y へのラベル γ の辺が存在し、 $w \in \gamma$ である
- x から z へラベル $\alpha \subseteq r, w, c$ の辺が存在する

を満たすとき、 G を次のように書き換えたグラフ G' を生成する。

- y から z へラベル α の辺を追加する

(c) Create

グラフ G に頂点 x が存在するとき、 G を次のように書き換えたグラフ G' を生成する。

- 新しい頂点 N を追加する
- x から N へラベル $\{r, w, c\}$ の辺を追加する

(d) Call

グラフ G に相異なる頂点 x, y, z が存在して条件

- x から z へのラベル γ の辺が存在し、 $c \in \gamma$ である
- x から y へラベル $\alpha \subseteq r, w, c$ の辺が存在する

を満たすとき、 G を次のように書き換えたグラフ G_0 を生成する。

- 新しい頂点 N を追加する
- N から z へラベル r の辺を追加する
- N から y へラベル α の辺を追加する

(e) Remove

グラフ G に相異なる頂点 x, y が存在して条件

- x から y へラベル $\alpha \subseteq r, w, c$ の辺が存在する

を満たすとき、 G を次のように書き換えたグラフ G_0 を生成する。

- x から y へのラベル α の辺を削除する

定理 1

p と q を保護グラフの頂点、 α をラベルとする。 p が q に対して α が可能であるための必要十分条件は、次の 2 つである。

(a) p と q が、グラフ G において連結であること。ここで、 p と q が 1 つ以上の辺で結ばれるとき、辺の向きやラベルにかかわらず、 p と q が連結であると呼ぶ。

(b) G の中に頂点 x と、 x から q へのラベル β の辺が存在し、次を満たすこと。

- $r \in \alpha$ ならば、 $\{r, c\} \cap \beta \neq \emptyset$

- $w \in \alpha$ ならば、 $w \in \beta$
- $c \in \alpha$ ならば、 $c \in \beta$

系 1

p が q に対して α 可能であることを、グラフの大きさに関して線形時間で判定するアルゴリズムが存在する。

系 2

p が q に対して α 可能であるならば、 p から q へラベル α の辺を、 p と q の間の経路に関して線形時間で追加するアルゴリズムが存在する。

4.2.2.2 主体と対象から構成されるモデル

実用的なシステムの多くでは、主体と対象は区別される。このようなシステムは、以下のような S-O take-grant システムによってモデル化される。

システムの保護状態を、次の要素からなる有限の有向グラフ（以下では、主体-対象保護グラフと呼ぶ）によってモデル化する。ここで、主体の集合と対象の集合は互いに素であるとする。

• 頂点

主体または対象。

• ラベル付きの有向辺

主体および対象が持つ許可。ラベルは権限を表し、 $\{r\}$ 、 $\{w\}$ 、 $\{r,w\}$ のいずれかである。

S-O take-grant システムの主要な書き換え規則は、次の 3 個である。

(a) Take

グラフ G に相異なる頂点 x 、 y 、 z が存在して条件

- x は主体、 y と z は主体または対象である
- x から y へのラベル r の辺が存在し、 $r \in \gamma$ である
- y から z へラベル $\alpha \subseteq \{r,w\}$ の辺が存在する

を満たすとき、 G を次のように書き換えたグラフ G' を生成する。

- x から z へラベル α の辺を追加する

(b) Grant

グラフ G に相異なる頂点 x 、 y 、 z が存在して条件

- x は主体、 y と z は主体または対象である
- x から y へのラベル r の辺が存在し、 $w \in r$ である
- x から z へラベル $\alpha \subseteq \{r,w\}$ の辺が存在する

を満たすとき、 G を次のように書き換えたグラフ G' を生成する。

- y から z へラベル α の辺を追加する

(c) Create

グラフ G に頂点 x が存在して条件

- x が主体である

を満たすとき、 G を次のように書き換えたグラフ G' を生成する。

- 新しい主体の頂点 N を追加する
- x から N へラベル $\{r, w\}$ の辺を追加する

定理 2

G を主体-対象保護グラフとし、 p_0 と q_0 を主体を表す頂点とする。また、ある主体から q_0 へラベル $\alpha \in r, w$ の辺があるものとする。 p_0 が q_0 に対して α 可能であるための必要十分条件は、ブロックの列

$$B_1, \dots, B_m$$

が存在し、 p_0 が B_1 に、 q_0 が B_m に含まれるとともに、 $i = 1, \dots, m - 1$ に対して B_i から B_{i+1} へのブリッジが存在すること。

ここで、ブロックは主体である頂点の直接連結（辺で直接結ばれる）関係による閉包。ブリッジは、一方のブロック内の頂点 p と他のブロック内の頂点 q の間の経路

$$p, x_1, \dots, x_k, q$$

であり、辺の並びが次のいずれかであるもの。

- (a) $p \xrightarrow{r} x_1 \xrightarrow{r} \dots \xrightarrow{r} x_k \xrightarrow{r} q$
- (b) $p \xleftarrow{r} x_1 \xleftarrow{r} \dots \xleftarrow{r} x_k \xleftarrow{r} q$
- (c) $p \xrightarrow{r} x_1 \xrightarrow{r} \dots \xrightarrow{r} x_i \xrightarrow{w} x_{i+1} \xleftarrow{r} \dots \xleftarrow{r} x_k \xleftarrow{r} q$
- (d) $p \xleftarrow{w} x_1 \xleftarrow{r} \dots \xleftarrow{r} x_k \xleftarrow{r} q$
- (e) $p \xrightarrow{r} x_1 \xrightarrow{r} \dots \xrightarrow{r} x_k \xrightarrow{w} q$
- (f) $p \xrightarrow{r} x_1 \xrightarrow{r} \dots \xrightarrow{r} x_i \xleftarrow{w} x_{i+1} \xleftarrow{r} \dots \xleftarrow{r} x_k \xleftarrow{r} q$
- (g) $p \xleftarrow{w} x_1 \xleftarrow{r} \dots \xleftarrow{r} x_k \xleftarrow{r} q$
- (h) $p \xrightarrow{r} x_1 \xrightarrow{r} \dots \xrightarrow{r} x_k \xleftarrow{w} q$

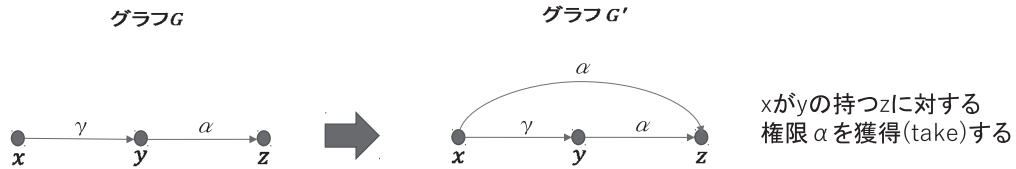
ただし、 $y \xrightarrow{\alpha} z$ は頂点 y から z へのラベルに α を含む辺があることを、 $y \xleftarrow{\alpha} z$ は頂点 z から y へのラベルに α を含む辺があることを示す。また、 $y \xrightarrow{\alpha} \dots \xrightarrow{\alpha} z$ は頂点 y から z へラベルに α を含む辺の 0 個以上の並びがあることを示し、辺が 0 個のときは $y = z$ である。同様に、 $y \xleftarrow{\alpha} \dots \xleftarrow{\alpha} z$ は頂点 z から y へラベルに α を含む辺の 0 個以上の並びがあることを示し、辺が 0 個のときは $y = z$ である。

系 3

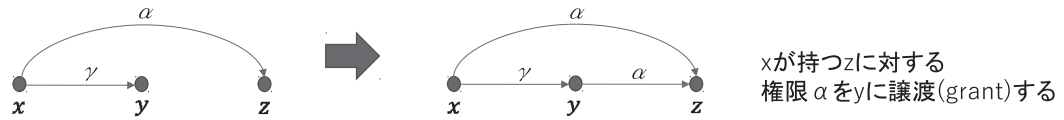
S-O take-grant システムでは、 p が q に対して α 可能であるかどうかを決定するアルゴリズムが存在する。^[17]

take, grant, create の書き換え規則を図 21 に示す。

▪Take



▪Grant



▪Create

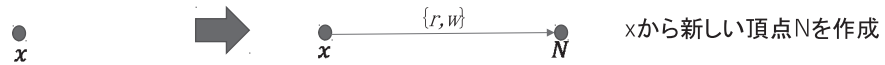


図 21 take-grant システムの主要な書き換え規則

4.2.3 スマートプロパティ

スマートプロパティとは、ある資産（プロパティ）を仮想通貨のようにコンピュータネットワーク上で流通できるようにしたものである。プロパティ情報を含むコインを生成してそのコインをブロックチェーン上で流通でき、ブロックチェーン技術の高い改ざん耐性で資産情報を保護することができる。スマートプロパティではプロパティに応じて任意の実行を実行するために、プログラムコードをブロックチェーンに埋め込み任意の実行を実行するため、新たなブロックチェーンが必要とされ新たなプラットフォームが登場している。その代表例が Ethereum である。

スマートプロパティを行使するということはスマートコントラクトを実行するということを意味する。なお、スマートプロパティは車などの実物の所有権だけでなく、著作権や特許などの目に見えないものも含まれると考えられる。^{[21][22]}

4.3 提案手法

情報カプセルは暗号化されたコンテンツと管理のためのメタデータ、所有者のエージェント、情報カプセルのエージェント、ユーザ側のアクセスコントロールエージェントからなる。アクセスコントロールエージェントからのコンテンツの利用要求を情報カプセルのエージェントが処理するメタデータの情報に基づいて利用が正当と評価できれば許可を出す。メタデータの情報だけでは決定不能な場合は所有者のエージェントとネゴシエーションする。これによりコンテンツの二次利用の際の自律的な著作権解決を可能にする。このシステムは個人情報の利用の管理にも適応でき、個人情報の利用目的に即した利用の保証や正当性の検査を行うことが可能となる。またコンテンツが編集・利用された結果を権利者が知るために、プライバシーを保ちつつ流通をトレースすることが可能となる。

そして、情報カプセルは、権利者に関する情報、アクセス権や利用条件などのメタデータをエージェント同士で行う必要があるため、エージェントの現在の所有者が誰であることを記録し常に更新しておく必要がある。そこで、メタデータをブロックチェーンに記録することでエージェントのアクセスが

容易になる。さらに、ブロックチェーンを応用したスマートプロパティおよびスマートコントラクトと連携させることでコンテンツに関する権利や利用条件などを信頼できる第三者なしで保証することが可能となる。また、情報カプセルを構成するエージェントをコントラクトコードと連携することによりエージェントの動作を保証する。従来提案されているスマートプロパティは、権利者の移転を記述することは可能であったが、コンテンツに対するアクセス権は従来の DRM の枠組みに依存していた。コントラクトコードと従来の DRM の枠組みを連携させることでスマートプロパティによる権利管理を確実なものとすることができる。

したがって、本研究では情報カプセルとブロックチェーンを用いたデジタルコンテンツの流通を提案する。

1. デジタルコンテンツをカプセル化する（情報カプセル）
2. Take-Grant モデルを情報カプセルのメタデータであるアクセス権の委譲に利用
3. 情報カプセルをブロックチェーンネットワークで流通させる

提案するデジタルコンテンツ流通のイメージを図 22 に示す。

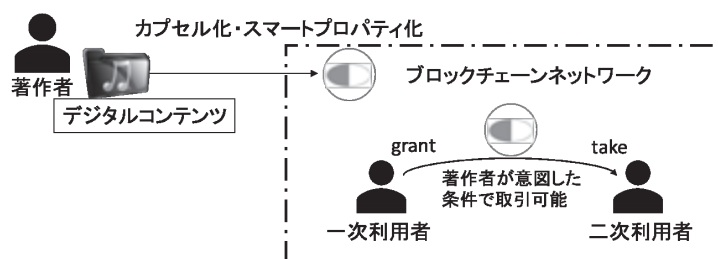


図 22 提案するデジタルコンテンツ流通のイメージ

4.3.1 提案システムの概要

プログラムコード埋め込み型ブロックチェーンを情報カプセルとみなし、Take-Grant による分析コードを記述する。デジタルコンテンツの利用に take 権と grant 権を設定し、利用条件を照会し満たしている場合のみコンテンツを利用可能とすれば、不正利用が不可能になると考える。今回提案するブロックチェーンのイメージを図 23 に示す。

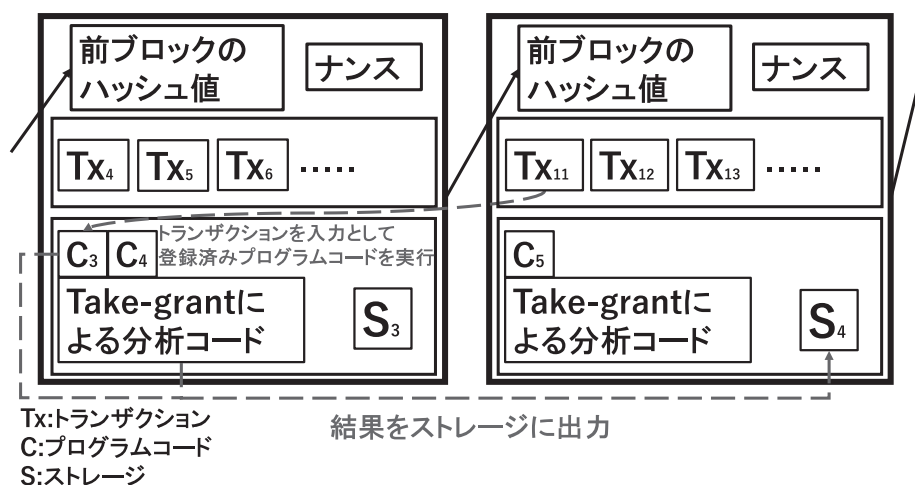


図 23 提案するブロックチェーンのイメージ

4.3.2 コントラクトコードの設計

スマートコントラクトを実行するためのプログラムコードをコントラクトコードと呼び、その設計をする。コントラクトコード実行の流れを図 24 に示す。



図 24 コントラクトコード実行の流れ

1. {S} は C を実行できるという権限を持つ。
2. {S} は C により結果をストレージに書き込む権限を持つ。
3. 権限はコード内の function で表され、これについて Take 権、Grant 権を設定し分析する。

次に利用者の Take 権とコンテンツの Grant 権が照会できた場合にコンテンツの利用が可能となるコントラクトコードの設計をする。

1. 著作者が Grant 権や Take 権の付与・譲渡の条件を設定したコントラクトを作成。
- 2.1 次利用者がコントラクトに対して Take 権を申請し、Take 権を獲得する。
- 3.1 次利用者の Take 権とコンテンツの Grant 権を照会する。
4. 照会に成功したら 2 次利用者のデジタルコンテンツの利用が可能になる。
- 5.1 次利用者が著作者が設定した条件で 2 次利用者に Take 権を譲渡。
- 6.2 次利用者の Take 権とコンテンツの Grant 権を照会する。
7. 照会に成功したら 2 次利用者のデジタルコンテンツの利用が可能になる。

以上の流れを図 25、図 26 に示す。

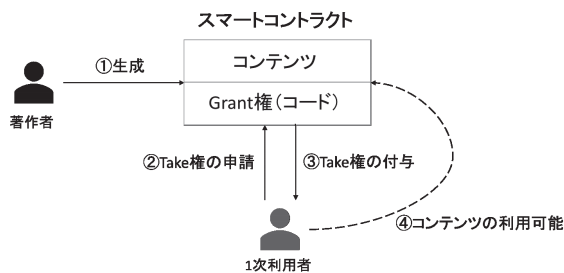


図 25 コントラクトコードの設計 (1)

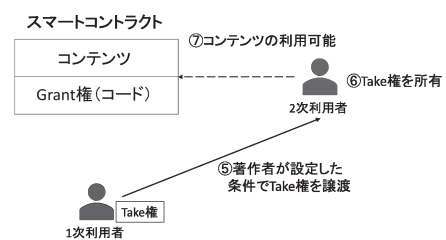


図 26 コントラクトコードの設計 (2)

4.4 結論

本研究では、情報カプセルとプログラム埋め込み型ブロックチェーンを用いることで、著作者の権利を保護しつつデジタルコンテンツの利用者同士でのやり取りを実質的に可能とした。さらにセキュリティモデルとして Take-Grant モデルを採用し、スマートコントラクトを用いることで、利用権の移転を Take-Grant モデルで記述して分析・修正を可能としてブロックチェーンにより権利の健全性が示せるデジタルコンテンツの流通システムを提案した。これにより従来よりもデジタルコンテンツ

の著作権者の権利の保護とシステムのセキュリティを確立できると考える。

今後の課題として、消費コストを抑え、スマートコントラクトにおけるセキュリティを考慮したコード設計や実際のデジタルコンテンツとどうやって連携させるかを考える必要がある。また、今回のスマートコントラクト開発環境では Ether しか仮想通貨を扱えない、Solidity といった専用言語しか扱えない、コントラクト生成時などにコストがかかってしまうなどの問題があるため、どの仮想通貨や言語にも対応しコントラクト生成コストが少ないスマートコントラクト開発環境の導入などが挙げられる。

5 LDA を用いた非文字資料検索法

5.1 まえがき

非文字資料研究センターの非文字資料に関連する研究課程・研究成果は『非文字資料研究』（旧年報）として年 2 回刊行されている。この『非文字資料研究』（旧年報）は神奈川大学日本常民文化研究所非文字資料研究センターの Web で公開されており誰でも閲覧できる。この資料を本論文では非文字年報と呼ぶ。非文字年報は研究分野ごとに分類されておらず専門的な単語も出てくる。よって専門的な知識を持たないユーザがこれらの資料を適切に活用するのは難しい。また非文字年報とは異なる別の資料との横断検索を考えた場合、分野や文書の体裁、資料の分類の違いによって、従来の全文検索では合理的ではない。テキストデータを解析し膨大なデータを扱いやすくする試みに、LDA トピックモデルを用いてカウンセリングデータの文書分類を研究しているものがある。カウンセリングをコンピュータで行うことによってプライバシー保護やコストの低下が可能となる。カウンセリング利用者の発言を理解する方法として LDA を用いたカテゴリ分類での文書分類の有効性が証明されている^[31]。画像から特徴量を抽出してベクトル量子化を用いて BoW 表現できる形にし、文書と画像から BoW 表現により統合した単語集合を抽出し、これにトピック分布を適用することによって文書と画像間の相互的な検索を可能とし、その有用性を示している^[32]。

LDA で文書を構成するトピック分布とそれぞれのトピックに属する単語の分布を出力することができ、各文書のトピック分布を比較することによって文書間の潜在的な繋がりを確率的に求めることができる^[33]。ここで文書を構成するトピックの分布とトピックを構成する単語の分布にベイズの定理を適用すると単語に疑似的なトピック分布を付与することができる。これによってトピック分布を介して単語と文書間の繋がりを確率的に求めることが可能となる。本研究では資料検索改善のため、非文字年報を対象として文書解析にトピックモデルを用いて、出力された単語と文書のトピック分布から類似の文書と単語を導出しデータベースを作成する。ユーザが閲覧した文書情報から、次にユーザが閲覧すべき文書を提示するシステムの構築を目的とする。

5.2 提案手法

5.2.1 提案システムの概念

本論文において、「非文字とはトピックモデルにおいて使用される LDA (Latent Dirichlet Allocation: Blei and McAuliffe, “Supervised Topic Models”, in Advances in Neural Information Processing Systems 20

(Proc. NIPS), 2008.) で定義される潜在的確率変数である」と定義する。

本研究のオリジナリティは、大雑把に言えば LDA で計算される事後確率によって観測可能となる潜在的確率変数を、すでにそれは観測可能になっているのだからもはや潜在的ではなく顕在化した確率変数と見なし、その確率分布をベイズの定理から計算しよう、というものである。以下にその概念を数学的に示す。

LDA はトピックと見なされる単語によって表現される潜在的確率変数 z_n がテキストに観測される単語 ω_n を生成する、と見做す確率的生成モデルである。即ち、潜在的確率変数 z_n が単語確率変数 ω_n を生成するとして $p(\omega_n | z_n)$ を事後確率と見なし、ベイズ推定する確率モデルである。

即ちベイズの定理より近似的に：

$$p(\omega_n | z_n) \propto p(z_n | \omega_n) \quad (\text{但し、} \propto \text{ は比例})$$

が成立する。トピックモデルに於けるトピックとは潜在的確率変数 z_n 、言い換えればトピックとしての単語を示す確率変数である。しかし、LDA に於いて正確には、上記比例式右辺 $p(z_n | \omega_n)$ を尤度とすることはない。上記比例式はあくまで概念的近似表現である。図 27 に示される LDA では、確率分布としてディリクレ分布が用いられ、潜在的確率変数が確率測度空間に位置すると定義し、あたかも確率の確率を表現しているかのように見せる理論である。

さて、ここで事後確率 $p(\omega_n | z_n)$ が計算されたということは、トピック z_n が計算されているということである。即ち潜在的確率変数 z_n が顕在化した観測単語確率変数 ζ_n として手に入った状況になったということである。そうであるならば、逆に上記比例式からベイズの定理より：

$$p(\zeta_n | \omega_n) \propto p(\omega_n | z_n)$$

が成立する。この計算は、既に確率変数 ω_n 、 z_n 、 ζ_n が観測されているので、ディリクレ分布を使用する必要はなく、それぞれの頻度分布を用いて計算すればよい。

本論文に於いては LDA の計算に Gensim package を使用した。Gensim には上記に示す原理的なディリクレ配分法に追加して、単語 ω_n が含まれているテキスト名との関係も計算可能になっている。単語 ω_n とトピック、そしてそれらを包含するテキスト集合との関係に関しては本論文の以下の節において説明する。提案するシステムを図 27 に示す。

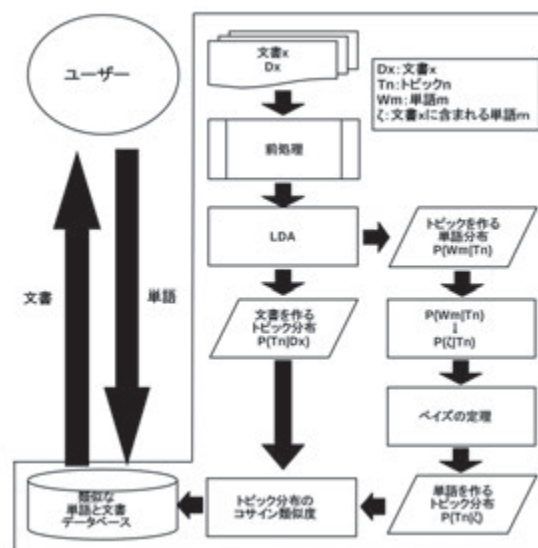


図 27 提案システム図

5.2.2 前処理

あらかじめ文書を用意する。今回の解析対象は神奈川大学非文字資料研究センターから刊行されている非文字資料に関する研究成果、共同研究の成果を論文として収録した『非文字資料研究』(旧年報)のうち日本語のもの(10MB)を使用する。^[30]

《形態素解析》

トピックモデルを作成するためには文書を形態素(言語で意味を持つ最小単位)に分解する必要がある。^[35]日本語の文は英語の文と異なり単語と単語の区切りが少ないため、あらかじめ用意した単語とその品詞について記述された辞書をもとに日本語の文法的に自然な繋がりになるようにする。^[36]また、本実験ではトピックモデルに使用する単語の品詞を指定するため、分かち書きされた単語の品詞を判別できるツールが望ましい。日本語の形態素解析ツールには「ChaSen」や「MeCab」などがあるが、^{[37][38][39]}今回はインストールが容易で動作が高速な「MeCab」を採用した。また形態素解析に用いる辞書は固有表現に強い「mecab-ipadic-NEologd」^[40]を使用する。

《学習用コーパス作成》

文書を形態素に分解したものの中で必要な品詞(本実験では名詞・動詞・形容詞)に絞って記録する。このとき、除去しきれない不要な単語や記号を別途で除去する。

Gensim に用意されている corpora.Dictionary という単語と ID のマッピングを行うツールを使い単語を ID で管理しする。corpora.Dictionary に用意されているフィルタ機能を利用して形態素解析や単語除去では取り除くことのできなかった、“極端に出現頻度が高い単語”や“極端に出現数の低い単語”のようなトピックとしての特徴が出にくいものを取り除き、Document-term matrix(単語の出現頻度をドキュメントごとに記述したもの)形式でコーパスを作成する。

5.2.3 LDA

作成されたコーパスを用いて LDA を実行する。Gensim の LDA では事前にトピック数を定める必要があるが、LDA の出力であるトピックを構成する単語の分布がトピックごとにわけられていると判断できたトピック数 20 に定めた。

LDA の出力のうち、文書を構成するトピックの分布を示すものを $P(T_n|D_x)$ 、トピックを構成する単語の分布を $P(W_m|T_n)$ とする。

・ T_n : トピック n ・ D_x : 文書 x ・ W_m : 文書 m

Gensim の LDA の出力の一部を図 28、29 に示す。

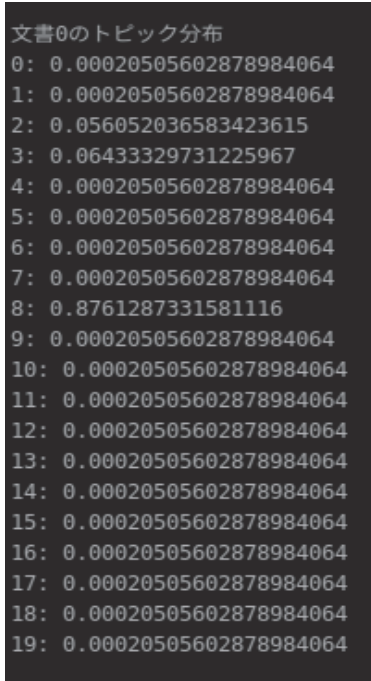


図 28 文書を構成するトピック分布



図 29 トピックを構成する単語分布

5.2.4 単語を構成するトピック分布の算出

《単語の分布を文書ごとの単語の分布に変換》

トピックを構成する単語分布 $P(W_m|T_n)$ の示す“単語”は“全文書で出現する単語”を示している。本実験では“特定の文書に属する単語”を示す必要があるため以下の計算を実行する。

- ・ ζ は文書 x に属する単語 m を示す。

$$P(\zeta | T_n) = P(W_m | T_n) * P(T_n | D_x)$$

トピックを構成する（どの文書に属しているかに判別されている）単語分布 $P(\zeta | T_n)$ が求まる。

《ベイズの定理の適用》

本実験では $P(T_n | \zeta)$ を使用するため、 $P(\zeta | T_n)$ にベイズの定理を適用する。

$$P(T_n | \zeta) = \frac{P(\zeta | T_n)P(T_n)}{P(\zeta)}$$

未知の値は以下の式で算出する。

$$P(T_n) = \frac{\sum_{x=0}^a P(T_n | D_x)}{a} \dots (a: \text{文書の総数})$$

$$P(\zeta) = \sum_{m=0}^b P(\zeta | T_n) * P(T_n) \dots (b: \text{トピックの総数})$$

5.2.5 データベースの作成

ドキュメントを構成するトピック分布 $P(T_n | D_x)$ と、（属する文書が判別された）単語を構成するトピック分布 $P(T_n | \zeta)$ から単語と文書の類似度をそれぞれのトピック分布のコサイン類似度として算出しデータベースに記録する。

5.2.6 実装環境

ユーザが閲覧した文書を構成するトピックを表示し、その中からユーザ自身が必要とするトピックを選択し、閲覧した文書と類似な文書の中でも、ユーザが必要としているトピックを多く含む文書を提示する。実装環境 :Ubuntu 18.04.4 LTS, プロセッサ :Intel Core i7-7567U CPU @ 3.50GHz × 4, グラフィックス :Intel Iris Plus Graphics 650 (Kaby Lake GT3e), メモリ :31.3 GiB, Python3.7.4, Anaconda 4.7.12, Gensim3.8.0

5.3 結果

作成した、単語と類似の文書とその類似度データベースの一部を図 30 に示す。また、トピックモデルに使用する品詞が名詞のみの場合の結果を図 31 に示す。

5.4 考察

今回の研究ではデータベースの作成までに留まってしまった。よって作成されたデータベースについて考察する。

図 30 の「中国・韓国の旧日本租界」の文書に含まれる単語の多くは「租界研究の現状と展望」と類似しているという傾向が見られる。実際に双方の文書の全文を読んで比較すると類似した文書であることが確認できた。このことから”文書”に対して類似した文書を抽出することは実現している。

	doc_name	oc_pa	word_name	sim_doc1	oc_pa	sim_pro1	sim_doc2	oc_pa	sim_pro2
1	中国・韓国の旧日本租界…	ne…	10年	租界研究の現状と展望	ne…	0.9996…	模地絵図を読み解く ― …	ne…	0.9996…
2	中国・韓国の旧日本租界…	ne…	2年間	歴史民俗資料としてみる『風俗画報』の再…	ne…	0.9752…	中国・韓国の旧日本租界 …	ne…	0.9745…
3	中国・韓国の旧日本租界…	ne…	3年間	租界研究の現状と展望	ne…	0.9977…	模地絵図を読み解く ― …	ne…	0.9977…
4	中国・韓国の旧日本租界…	ne…	coe	歴史民俗資料としてみる『風俗画報』の再…	ne…	0.9752…	中国・韓国の旧日本租界 …	ne…	0.9726…
5	中国・韓国の旧日本租界…	ne…	お願い	租界研究の現状と展望	ne…	0.9999…	模地絵図を読み解く ― …	ne…	0.9999…
6	中国・韓国の旧日本租界…	ne…	もらい	中国・韓国の旧日本租界 ― 研究班の3…	ne…	0.9921…	租界研究の現状と展望	ne…	0.9830…
7	中国・韓国の旧日本租界…	ne…	もらう	中国・韓国の旧日本租界 ― 研究班の3…	ne…	0.9986…	租界研究の現状と展望	ne…	0.9929…
8	中国・韓国の旧日本租界…	ne…	キャンパス	歴史民俗資料としてみる『風俗画報』の再…	ne…	0.9746…	中国・韓国の旧日本租界 …	ne…	0.9701…
9	中国・韓国の旧日本租界…	ne…	コメント	租界研究の現状と展望	ne…	0.9993…	模地絵図を読み解く ― …	ne…	0.9993…
10	中国・韓国の旧日本租界…	ne…	シンボル	中国・韓国の旧日本租界 ― 研究班の3…	ne…	0.9993…	租界研究の現状と展望	ne…	0.9930…
11	中国・韓国の旧日本租界…	ne…	シンポジウム	中国・韓国の旧日本租界 ― 研究班の3…	ne…	0.9979…	租界研究の現状と展望	ne…	0.9964…
12	中国・韓国の旧日本租界…	ne…	セッション	租界研究の現状と展望	ne…	0.9995…	模地絵図を読み解く ― …	ne…	0.9995…
13	中国・韓国の旧日本租界…	ne…	タイトル	中国・韓国の旧日本租界 ― 研究班の3…	ne…	0.9979…	租界研究の現状と展望	ne…	0.9978…
14	中国・韓国の旧日本租界…	ne…	デジタル	サハリンにおける「カラフト」期の日本文…	ne…	0.9999…	民俗学確立以前における図…	ne…	0.9953…
15	中国・韓国の旧日本租界…	ne…	データベース	ネットワーク形成と中国非物質文化遺産研究	ne…	0.9369…	日本の敗戦後における旧唐…	ne…	0.8907…
16	中国・韓国の旧日本租界…	ne…	ネットワーク	歴史民俗資料としてみる『風俗画報』の再…	ne…	0.9707…	中国・韓国の旧日本租界 …	ne…	0.9539…
17	中国・韓国の旧日本租界…	ne…	プロパガンダ	中国・韓国の旧日本租界 ― 研究班の3…	ne…	0.9977…	租界研究の現状と展望	ne…	0.9956…
18	中国・韓国の旧日本租界…	ne…	ホテル	租界研究の現状と展望	ne…	0.9998…	模地絵図を読み解く ― …	ne…	0.9998…
19	中国・韓国の旧日本租界…	ne…	ポスター	近藤恒弘氏に天津日本租界での体験を聞く	ne…	0.9559…	東西にみる道祖神の現状	ne…	0.9204…
20	中国・韓国の旧日本租界…	ne…	上海	租界研究の現状と展望	ne…	0.9993…	模地絵図を読み解く ― …	ne…	0.9993…
21	中国・韓国の旧日本租界…	ne…	中国人	租界研究の現状と展望	ne…	0.9983…	模地絵図を読み解く ― …	ne…	0.9983…
22	中国・韓国の旧日本租界…	ne…	中国語	租界研究の現状と展望	ne…	0.9998…	模地絵図を読み解く ― …	ne…	0.9998…
23	中国・韓国の旧日本租界…	ne…	中華民国	租界研究の現状と展望	ne…	0.9984…	模地絵図を読み解く ― …	ne…	0.9984…
24	中国・韓国の旧日本租界…	ne…	仁川	租界研究の現状と展望	ne…	1.0	模地絵図を読み解く ― …	ne…	1.0
25	中国・韓国の旧日本租界…	ne…	企業	租界研究の現状と展望	ne…	0.9974…	模地絵図を読み解く ― …	ne…	0.9974…
26	中国・韓国の旧日本租界…	ne…	会場	歴史民俗資料としてみる『風俗画報』の再…	ne…	0.9690…	中国・韓国の旧日本租界 …	ne…	0.9682…
27	中国・韓国の旧日本租界…	ne…	会場	歴史民俗資料としてみる『風俗画報』の再…	ne…	0.9690…	中国・韓国の旧日本租界 …	ne…	0.9682…

図 30 単語と文書の類似度データベース

文書間のトピック分布類似度を計算した場合は当然同じ文書が最も類似しているものとして挙げられるが、単語の持つトピック分布を考慮した本実験では、同じ文書が最も類似度の高いものとして挙げられているものが少ない。よって、単語を構成するトピック分布に文書だけでなく単語の持つ意味の要素も含まれていることが考えられる。

「デジタル」、「データベース」、「ネットワーク」は主観評価で同じトピックに分類されているように見えるが、データベースでは類似した文書はそれぞれ別なものが当てはめられている。原因として考えられるのは学習に用いた文書の量の少なさによる学習不足である。全体の文書をみても記述の多い「中国人」「中国語」「中華民国」などの「中国」をトピックとしている単語と類似の文書は同じものが挙げられている。このことから本実験においては学習用の文書量が少なかったと考えられる。

図 31 はトピックモデルで扱う品詞が名詞のみの場合である。図 30 と比較して類似の文書として出てくる「租界研究の現状と展望」の割合が少ない。「租界研究の現状と展望」は「中国・韓国の旧日本租界」と類似した文書であると主観評価で確認しているため、この結果は不適切である。このことから、トピックモデルを作成する際に用いる品詞は「名詞」だけではなく「動詞」「形容詞」も追加したほうが精度が上がると考えられる。また文書の量の少ない今回の実験では、LDA を実行する際に使用する単語の数が増えることも使用する品詞が少ないと精度が落ちる原因につながる可能性がある。

これらの考察から本システムは、文書と単語から類似の文書を探すことにに関して一定の目安にはなるが、精度の面でもユーザビリティの面でも低いものと考えられる。

	doc_name	oc_par	word_name	sim_doc1	oc_par	sim_pro1	sim_doc2	oc_par	sim_pro2
1	中国・韓国の旧日本租界	ne	10年	チベット仏教寺院におけるバリン儀礼の比較研	ne	0.8919	戦後スペインと日本における学生運動のイメ	ne	0.8919
2	中国・韓国の旧日本租界	ne	2年間	チベット仏教寺院におけるバリン儀礼の比較研	ne	0.9765	戦後スペインと日本における学生運動のイメ	ne	0.9765
3	中国・韓国の旧日本租界	ne	3年間	中国・韓国の旧日本租界	ne	0.9708	教会大学と日中戦争「北平私立輔仁大学 案	ne	0.9549
4	中国・韓国の旧日本租界	ne	coe	日本人が見た上海イメージ	ne	0.9926	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9926
5	中国・韓国の旧日本租界	ne	お願い	日本人が見た上海イメージ	ne	0.9790	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9790
6	中国・韓国の旧日本租界	ne	まとめ	チベット仏教寺院におけるバリン儀礼の比較研	ne	0.9476	戦後スペインと日本における学生運動のイメ	ne	0.9476
7	中国・韓国の旧日本租界	ne	キャンパス	教会大学と日中戦争「北平私立輔仁大学 案	ne	0.9937	中国・韓国の旧日本租界	ne	0.8870
8	中国・韓国の旧日本租界	ne	コメント	中国・韓国の旧日本租界	ne	0.9716	教会大学と日中戦争「北平私立輔仁大学 案	ne	0.8925
9	中国・韓国の旧日本租界	ne	シンボル	福州神社跡地追加報告—創建当時に関する関	ne	0.7603	『海外神社跡地のその後』共同研究	ne	0.7538
10	中国・韓国の旧日本租界	ne	シンポジウム	日本人が見た上海イメージ	ne	0.9942	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9942
11	中国・韓国の旧日本租界	ne	セッション	日本人が見た上海イメージ	ne	0.9999	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9999
12	中国・韓国の旧日本租界	ne	タイトル	日本人が見た上海イメージ	ne	0.9951	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9951
13	中国・韓国の旧日本租界	ne	デジタル	日本人が見た上海イメージ	ne	0.9999	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9999
14	中国・韓国の旧日本租界	ne	データベース	『戦時下日本の大衆メディア研究』	ne	0.9588	日本人が見た上海イメージ	ne	0.9496
15	中国・韓国の旧日本租界	ne	ネットワーク	日本人が見た上海イメージ	ne	0.9809	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9809
16	中国・韓国の旧日本租界	ne	プロパガンダ	日本人が見た上海イメージ	ne	0.9999	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9999
17	中国・韓国の旧日本租界	ne	ホテル	東アジアの租界とメディア空間	ne	0.9235	侵略神社跡地調査・ソウル	ne	0.8538
18	中国・韓国の旧日本租界	ne	ポスター	日系チャプスイレストランにおけるフォーチュ	ne	0.9998	旧樺太時代の神社について—併せて北方領土の	ne	0.9998
19	中国・韓国の旧日本租界	ne	上海	日本人が見た上海イメージ	ne	0.9995	監仙信仰と邪疫治療—湖北省丹江口市山村社会	ne	0.9995
20	中国・韓国の旧日本租界	ne	中国人	教会大学と日中戦争「北平私立輔仁大学 案	ne	0.9738	中国・韓国の旧日本租界	ne	0.9273
21	中国・韓国の旧日本租界	ne	中国語	教会大学と日中戦争「北平私立輔仁大学 案	ne	0.9852	中国・韓国の旧日本租界	ne	0.9463
22	中国・韓国の旧日本租界	ne	中華民国	中国・韓国の旧日本租界	ne	0.9109	教会大学と日中戦争「北平私立輔仁大学 案	ne	0.9089
23	中国・韓国の旧日本租界	ne	仁川	『戦時下日本の大衆メディア研究』	ne	0.9704	日本人が見た上海イメージ	ne	0.9683
24	中国・韓国の旧日本租界	ne	企業	東アジアの租界とメディア空間	ne	0.8591	チベット仏教寺院におけるバリン儀礼の比較研	ne	0.7898
25	中国・韓国の旧日本租界	ne	会場	チベット仏教寺院におけるバリン儀礼の比較研	ne	0.9827	戦後スペインと日本における学生運動のイメ	ne	0.9827
26	中国・韓国の旧日本租界	ne	会議	教会大学と日中戦争「北平私立輔仁大学 案	ne	0.9872	中国・韓国の旧日本租界	ne	0.8737

図 31 名詞のみでトピックモデルを作成した場合のデータベース

5.5 今後の課題

今回の実験でのシステムの評価は主観評価に頼ったものだったが、今後は客観に基づく評価をするためのデータや評価指標などを調べたい。

自然言語解析では機械学習の前処理が結果に大きく影響するので、より精度の高い機械学習を行うことのできる前処理の検討が必要。

今回の実験では単語のもつトピック分布に着目したが、今後は文書の持つトピック分布に着目して、ユーザが興味のあるトピックとそうでないトピックの重みを調整して類似の文書を提示するシステムについて検討したい。

6 潜在的な text に対する倫理的考察

クラウドに於ける情報セキュリティのあるべき「必要条件」とは何か？この問いに対し、まず、システムとして人工知能と情報理論に着目した。次に、ウィトゲンシュタインの「言語ゲーム」およびジャン・リュック＝ナンシーの「共同体（コルプス）」の哲学的概念を手掛かりとし、これらの概念を、仏教が示す「自利利他」あるいは世阿弥の「離見の見」から俯瞰した。その結果、上記必要条件は＜私＞と他者の身体性の視座から考える必要がある、という条件に行き着いた。この必要条件によれば、クラウドの情報セキュリティは、論理的な存在論に基づくセキュリティモデルのみによって論ぜられるべきではなく、クラウドに参画する人々の言説が確率的存在者として示され、情報理論的評価が実施されるシステムが必要である事が示される。言い換えれば、クラウドに於ける情報漏洩、情報改ざんの分析制御のためには、人工知能によるベイズ推論とそれに基づく存在論的情報フィルタを研究開発する方向性が示されたと言える。

今や、クラウドと AI は不可分のシステムである。世界の記述に於いて、世界は確率モデルによって確率測度空間に表れる。即ち、事象を確率測度空間に於いて Latent Dirichlet Allocation (LDA) 等で記述しベイズ推定する。AI が確率測度空間で何を示すのかは事象の確率的モデルによって示される。更に、世界の記述に於いてもう 1 つ必要なアイテム：学習 text がある。確率モデルの学習パラメータは入力される学習 text により決定される。即ち、AI に於ける世界の記述の限界は、AI の入力 text の選択如何で決まる。そして、この入力 text は人間の営みそのものを示している。したがって AI に於いては、人間が関与する限り人間の振る舞い次第でシステムの振る舞いの善し悪しが決まることに留意しなければならない。この意味で、インターネットから入手可能な機械学習パッケージを社会装置としてただ利用するだけでは、ソクラテスが言う「ただ生きるのではなく、善く生きること」という倫理的な振る舞いは達成されない。社会の中で「善く生きる事」が情報セキュリティと AI の設計に反映されるためには、そもそも「善く生きる事」が何を意味するかを考え設計する振る舞いを継続しなければならない。「善く生きるための必要条件」は：

- ・「分有的なロゴスの領野に属する振る舞い」、或いは「差異として捉えられるべき振る舞いである」、とする。
- ・言い換えれば、原理として命題化される行為は善く生きるための必要十分条件ではない。
- ・更には差異で捉える世界を確率測度空間に於いて示される振る舞いと見なす。確率測度は、確率

変数という存在者の存在を示す積分可能な尺度である。

振る舞いは確率変数としての存在者を表しており、その存在として「あるべき姿」は確率分布の機械学習によって捉えられ、その評価は情報量によって示される。

しかしながら一方で社会システムは AI を使用し、ある時、ある場所に於ける「原理」にしたがって、事象に対して確率的ではない「決定」をしなければならない。私たちはこの「決定」を、私たちの「原理」「規範」「規則」にしたがってプログラムしている。更に、AI の基礎：「ベイズ推定」は「分有ロゴスの視点に於いて主観的な推定」である。何故ならベイズ推定に於いて入力される学習 text は人間が自然言語で記述した text、或いは人間によってある意図のもとに恣意的に収集されたデータだからである。この場合の「主観」はナンシーの言う「分有ロゴス」に類する概念である。一方、論理的「規則」という普遍性、或いは数学的普遍性は客観的であるとされる。しかし、その事態と、個々の<私>が善く生きる事、は互いに不可分な必要十分条件ではない。なぜなら<私>は「個々の主観」という object ではないのである。そうだとするならば、<私>の身体はどのような「はたらき」としてこの一件に関わってくるのであろうか？私たちは次の命題を考えなければならないのである。：

- ・身体を伴う<私>は主観と客観の狭間でどの様にして善く生きるべきだろうか？

この命題をナンシーの分有ロゴスと確率的存在論から解釈すれば次のようになるであろう：分有ロゴスに於ける特異な声として出現する<私>と言う確率変数、即ち確率的存在者は、その分有ロゴスを示している言語ゲームを如何にして位置付ければ良いか？まず、ロゴスと言語ゲームの間に循環があるとする。即ち：

(1) 近代に於ける自己は共同体から解放され、論理的「理論」に基づく真理の探究が個々に可能になる一方で、

(2) しかしながら、生命体としての自己は共同体（コルプス）としての他者との絆を断ち切れ、「閉論理空間」という孤独な「言語ゲーム」に投げ入れられる。取り残された身体としての<私>は、身体性の必然として他者との繋がりを不安の内に「言語ゲーム」を介して希求する。これを契機として、特異な声としての<私>は共同体としての現れを「言語ゲーム」が示す確率測度空間に於いて見ることになる。

この循環 (1)(2) に対して更に、自利利他という視座に於いて身体性を導入し次の様に転回させる。即ち：

- ・ここで示す循環という状態を伴った作用は、身体性～として見れば「あたかも同時にある」ということである。
- ・「身体性～としてあたかも同時にある」、を言い換えれば、「論理学的存在者と確率測度的存在者は身体性に於いて分有される」である。
- ・そしてこの場合、「身体とはかくのごとき者であるとして自覚し、2つの存在者が存在者として存在するあり方を分有のうちに実践する振る舞い」が倫理の必要条件である。
- ・かくのごとく、ナンシーの分有ロゴスのシステムとウィトゲンシュタインの言語ゲームのシステムは身体性に於いて倫理的に統合される。
- ・そして、身体性に於いて同時に存在する、という事象を示す必要条件が確率測度空間に於ける確率的存在者としての<私>と他者、である。

AI とは text に潜在的に潜む確率変数で示されるクラスターを学習するシステムである。したがって従来の同値関係の原理によってのみ基礎付けられる倫理は、ここでは排除されなければならない。ベイジアン機械学習に於ける text は、人の営みによって生成される。したがって text の分析に於いて、原理よりもむしろ人の営みが先立たねばならない。「善く生きるための設計」とはそのような振る舞いに裏打ちされる運動である。即ち、世界の限界を臨むために技術と社会そして倫理にとっての必要条件は：

- ・「分有ロゴスとしての潜在的な text をベイジアンモデルに機械学習させること」である。
- ・或いは、「text に於ける潜在的なクラスター（即ち確率変数）を効果的に分析する AI architecture を設計すること」である。

ナンシーは「死」に於いて、「＜私＞は＜私＞の「死」を語ることができない事」から、＜私＞の世界の限界を示す。同様に、ナンシーは他者についても、「他者が＜私＞を見る事により自己の振る舞いを思い悩むという状況が生じる」とする。これが「人の生」を集合論的側面からのみ評価すると言う方法に於ける「世界の限界」である。ウィトゲンシュタインとナンシーがそれぞれ異なる視点に於いて示す世界の限界は、世界に於いて、世界を覆う集合とその要素だけしか見ていないと言うことから生じる限界である。＜私＞の終焉、即ち＜私＞の死に於いて、＜私＞は＜私＞の終焉を自ら語ることができない。しかしながら他者の終焉に対しては他者に対し、或いは他者の近傍の人々に対して終焉を回避する世界を確率測度空間の事象として示すことが可能な場合がある。そしてその様な「別世界への回避」こそが、世界の限界に多少なりともあらがうために必要である。ここに於いて、共同体（コルプス）と、＜私＞の間の分有した存在の在り方として、その存在の中に於いて身体性を伴う特異な声の探究、即ち「生身の身体を媒介とする存在の在り方の探究」、が必要になってくるのである。

参考文献

- [1] 只見町インターネット・エコミュージアム（参照 2019-01-06）。
<http://www.himoji.jp/tadamiitem/index.html>.
- [2] 大介小松、08 福島県南会津郡只見町の民具のデータベース化とその問題点・地域情報学の構築：新しい知のイノベーションへの道、神奈川大学、2008-03-10。
- [3] 吉賀夏子、渡辺健次、只木進一、貴重書書誌の注記から抽出したメタデータによるオントロジー構築および書誌・美術関連 linked data と連携した検索システム構築（「デジタル図書館」ワークショップ第 45 回発表論文）デジタル図書館、No. 45、pp. 3－10. デジタル図書館編集委員会、Sep. 2013。
- [4] Rie Jimbo, Sumiko Miyata, Kazumitsu Matsuzawa, and Hirotsugu Kinoshita. Rdfization of database for folk implements and ontology-assisted information retrieval system. In 2015 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-TW 2015, pp. 490－491. Institute of Electrical and Electronics Engineers Inc., 2015.
- [5] 只見町教育委員会、会津只見の生産用具と仕事着コレクション、北日本印刷、2005。
- [6] 只見町史編さん委員会、会津只見の民具、北日本印刷株式会社、2002。
- [7] 神崎正英、セマンティック・ウェブのための RDF/OWL 入門、森北出版、2010。
- [8] 溝口理一郎、知の科学 オントロジー工学の理論と実践、オーム社、2012。
- [9] 来村徳信、オントロジーの普及と応用、オーム社、2012。
- [10] 竹崎あかね、朱成敏、法隆大輔、武田英明、吉田智一、農業 IT システム間のデータ連携を推進する農作業

- 基本オントロジーの構築、農研機構研究報告農業技術革新工学研究センター = Bulletin of the NARO, Agricultural Machinery, No. 1, pp. 13 – 21, Mar. 2018.
- [11] 五十嵐達治、遠藤直樹、川森雅仁、古原和邦、三瓶徹、中西康治、“ユビキタス時代の著作権管理技術：DRMとコンテンツ流通”、東京電機大学出版局、2006。
- [12] 廣幡忠尚、“二次利用を想定した情報カプセルの提案と実装”、神奈川大学、2015 年度卒業論文。
- [13] Ogawa Nakamura, Takahashi, “A Copyright Information Embedding Method using DCT for Digital Movies”, 1997 信学春全大, D – 11 – 47.
- [14] Yamanaka Kiyoshi, “Digital watermarking technologies and the application to copyright protection : Series Intellectual property in information society”, Vol.40, No.10, Jan.1998.
- [15] 加賀美千春、森賀邦広、塩野入理、桜井紀彦、“コンテンツ流通における自律管理を目的としたカプセル化コンテンツ Matryoshka ”、NTT サイバースペース、マルチメディア通信と分散処理 .9718、コンピュータセキュリティ .818、Mar.2000。
- [16] A. K. Jones, R. J. Lipton, L. Snyder, “A LinearTime Algorithm for Deciding Security”, Proceedings of the 17th Annual Symposium on the Foundations of Computer Science, pp.33 – 41, 1976.
- [17] 独立行政法人情報処理推進機構、“アクセス制御に関するセキュリティポリシーモデルの調査 報告書 Take-Grant モデル”、2004 情財第 736 号、pp.136 – 142、2005。
- [18] 根本修二郎、“デジタルコンテンツの転々流通を可能にするビットコインの応用”、神奈川大学、2016 年度卒業論文。
- [19] 中園和孝、“ブロックチェーンを用いた電子セキュリティーキーの提案”、神奈川大学、2017 年度卒業論文。
- [20] EthereumRevision628d6654, “Solidity”, <https://solidity.readthedocs.io/en/develop/>, 参照 2019。
- [21] “ユーザー独自通貨とスマートプロパティ Bitcoin 日本語情報サイト”、
<http://jpbitcoin.com/bitcoin2/userasset>, 2016。
- [22] 大橋盛徳、NTT サービスエボリューション研究所、“スマートプロパティの概要とコンテンツ管理への適用”、2016。
- [23] 株式会社野村総合研究所、“平成 29 年度商取引適正化・製品安全に係る事業（ブロックチェーン技術を活用した新たなコンテンツビジネスの可能性調査）成果報告書”、2018。
- [24] Volker Roth, “IT-Security Access Control: Take-Grant Model”, Institut für Informatik Freie Universität Berlin Sommersemester 2009.
- [25] 安田浩、安原隆一、“ポイント図解式コンテンツ流通教科書”、アスキー、pp269、2003。
- [26] NTT データ 先端技術株式会社、“ブロックチェーン Ethereum 入門 1”、
<http://www.intellilink.co.jp/article/column/ethereum01.html>, 2016。
- [27] 総務省 平成 29 年版情報通信白書、“第 2 部 基本データと政策動向”、
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc261830.html>, 2017。
- [28] 独立行政法人情報処理推進機構、“アクセス制御に関するセキュリティポリシーモデルの調査 報告書 HRU モデル”、2004 情財第 736 号、pp.124 – 135、2005。
- [29] M. Harrison, W. Ruzzo, J. Ullman, “Protection in Operating Systems”, Communications of the ACM Vol.19, No. 8, pp.461 – 471 1976。
- [30] 神奈川大学 日本常民文化研究所 非文字資料研究センター Web URL : <http://himoji.kanagawau.ac.jp/index.html>, 参照 2020。
- [31] 単壮、加藤昇平、“カウンセリングデータにおけるトピックモデルを用いた文書分類”、電気学会論文誌 C 136 巻 2 号 p. 226 – 232 ,2016。
- [32] 矢田一貴、木下宏揚、“トピックモデルを用いた潜在的関連性を考慮した情報検索”、神奈川大学 2017 年度卒業論文。

- [33] 貞光九月、福富崇博、山本幹雄、“トピック教師なしデータからのトピック依存評価表現モデルの獲得”、電子情報通信学会研究報告 .NLC、言語理解とコミュニケーション 107、pp.57–62、2008。
- [34] Tomoyuki Nagahash, Arihito Ihara, Hironobu Fujiyoshi, “Tendency of Image Local Features that are Effective for Discrimination by using Bag-of-Features in Object Category Recognition,”、情報処理学会、CVIM 169、pp.1–8、2009。
- [35] 首藤公昭、“文節構造モデルによる日本語の機械処理に関する研究”、福岡大学研究所報、No.45、1980。
- [36] 長尾真、“講座現代の堅固 7 言語の機械処理”、三省堂、1984 年。
- [37] 長谷部陽一郎、“Wikipedia 日本語版をコーパスとして用いた言語研究の手法”、Doshisha Studies in Language and Culture、2006。
- [38] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <https://taku910.github.io/mecab/>, 参照 .2020。
- [39] ChaSen- 形態素解析器、<https://chasen-legacy.osdn.jp>、参照 2020。
- [40] 佐藤敏紀、橋本泰一、奥村学、“単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討”、言語処理学会第 23 回年次大会発表論文集 (NLP2017), NLP2017-B6-1, 2017。
- [41] 伊藤潤一郎、“ロゴスの分有 (パルタージュ): ジャン＝リュック・ナンシー『声の分有』における共同体”、表象・メディア研究、No.4、pp.91–110、2014。
- [42] 澤田直 (著)、“ジャン＝リュック・ナンシー: 分有のためのエチュード (哲学の現代を読む)”、白水社、(2013)。
- [43] ジャン＝リュック・ナンシー (著)、加藤恵介 (翻訳); 声の分有、松籟社。
- [44] ジャン＝リュック・ナンシー (著)、西谷修 (翻訳)、安原伸一郎 (翻訳); 無為の共同体、以文社、2001。
- [45] ウィトゲンシュタイン (著)、野矢茂樹 (訳)、“論理哲学論考”、岩波文庫、2003。
- [46] ウィトゲンシュタイン、藤本隆志 (訳)、菅豊彦 (訳)、“ウィトゲンシュタイン全集 8 哲学探究”、大修館書店。
- [47] 野矢茂樹、“規則とアスペクト: 『哲学探究』第 II 部からの展開”、北海道大学文学部紀要、The annual reports on cultural science, 36(2): pp.95–135, 1988。
- [48] David M. Blei, Jon D. McAuliffe, “Supervised topic models”, Proceeding NIPS’07 Proceedings of the 20th International Conference on Neural Information Processing Systems, pp.121–128. 2007.
- [49] Daniel Ramage, David Hall, Ramesh Nallapati and Christopher D. Manning, “Labeled LDA: A supervised topic model for credit attribution in multi- labeled corpora”, Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 248256, Singapore, August (2009).
- [50] Rajat Panda, Ankit Pensia, Nikhil Mehta, Mingyuan Zhou, Piyush Rai, “Deep Topic Models for Multi-label Learning, Proceedings of Machine Learning Research”, PMLR 89:2849–2857, 2019.
- [51] 森住哲也、“『離見の見』に於いて他者とは何か: 『離見の見』の構造”、電子情報通信学会、信学技報、vol. 116, no. 234, SITE2016-37, pp. 27–30, Oct.2016。
- [52] 森住哲也、“論理学的存在者から見る確率的存在者の倫理とは何か”、信学技報、vol. 117, no. 471, SITE2017-76, pp. 213–219, 3 月, (2018)。
- [53] 森住哲也、木下宏揚、“確率測度空間に於いて脱構築装置を内在するアクセス制御について”、信学技報、vol. 118, no. 152, SITE2018-28, pp. 281–287, 7 月, (2018)。
- [54] 森住哲也、“確率測度的テキストと決定論的テキストを循環させる逆強化学習システムの倫理とは何か”、信学技報、vol.118,no.345, SITE2018-62, 2018-12-06, pp.17–23, 12 月, (2018)。
- [55] 中谷憲・森住哲也・木下宏揚、“ベイジアンモデルによる情報漏えい分析のための機械学習”、電子情報通信学会ソサイエティ大会、A-12-1、9 月 12 日 (2018)。
- [56] プラトン (著)、久保勉 (翻訳)、“ソクラテスの弁明・クリトン”、岩波書店 1964 年。

7 今後の課題と展望

7.1 自己点検・評価

第三期共同研究では、インターネットエコミュージアムや只見町に開設予定の民俗博物館において必要な、データマイニングやデータの入力や検索に適したユーザインタフェースなどの基盤技術を開発することが目的であった。

知識とサービス、物の流通と価値交換

非文字資料を研究者間および一般ユーザと知識、サービス、資料をやり取りするために、ゲーム理論によりモデル化と解析および、自律分散的に事象や価値の移転を記録するブロックチェーンの技術を用いる計画であった。ブロックチェーンの利用に関しては、所有権の移転を保証するスマートプロパティや契約の履行を保証するスマートコントラクトを応用して著作物としてのコンテンツの流通を管理する手法などを提案できたので、要素技術については概ね目標を達成できた。サービスの流通については具体的な着手はできなかったので次期共同研究で行う予定である。一方、ゲーム理論に基づくモデル化は、ほとんど進展はしなかったので、次期共同研究で引き続き行う予定である。

知識とサービスの検索とマイニング

非文字資料のデータベースや研究者が資料を検索する際に、作業の流れであるコンテキストの一面に着目して情報の類似度などに基づいてユーザに対して最適な情報を提示することを目的としていた。情報検索では、意味解析において有望なトピックモデルを文書中に含まれるテキストと画像の両方に適用し、これを組み合わせることで潜在的データを抽出する手法、木構造のテキスト検索を行う際に非文字用のオントロジーに基づいた語と語の自己相互情報量による関連性の抽出、トピックモデルに基づく文書間の関連性の解析などを行うことができたので概ね目標は達成できた。次期共同研究では、これらを実装した実用レベルの検索システムを構築し、実データでの有効性を検証していく必要がある。サービスの検索については具体的な着手はできなかったので次期共同研究で行う予定である。

個人情報や重要情報、著作権の管理

情報保護や著作権管理にブロックチェーンや確率的な解析を行ったり、ブロックチェーンと組み合わせた新しい電子透かしシステムを応用することを目的とした。個人情報と重要情報の保護に関しては、推論攻撃の防止のために、文書レベルの確率的な関連性のトピックモデルによる解析、推論規則獲得のための Linked Open Data の活用、ブロックチェーンによるアクセス履歴の管理をベースとする推論攻撃の解析やスマートコントラクトを応用したアクセス制御など要素技術の提案を行った。電子透かしについては、ブロックチェーンに基づく電子透かしにより、従来問題となっていた透かし情報のデータサイズの拡張や多重電子透かしの管理の問題を解決し、このシステムに必要な不可欠な機械学習に基づく新しい知覚ハッシュを提案した。また、機械学習に基づく新しいファイアウォールの構築法を提案した。よって、目標は達成されたと考えられる。次期共同研究では、これらの成果をさらに発展させていく予定である。

潜在的な text のパターンを AI により健在化し、分類評価する研究 (パターンランゲージの研究)

この研究は、テーマ：「知識とサービスの検索とマイニング」および、テーマ：「個人情報や重要情報、著作権の管理」にとって、それらを概念設計する場合の必要条件としての「理念」を考察することに貢献した。その哲学的理念を示す成果は、技術研究報告という論文で既に発表済みである。具体的な「パターンランゲージ」は今だ実現するに至っていないが、その設計理念とビジョンは、より具体的な非文字関連の他の上記テーマを考察する形で関わりあうことにより、煮詰まりつつあると言える。

7.2 第5期事業計画

第4期の研究成果を踏まえて、「非文字資料の流通過程における諸問題を解決するための機械学習やブロックチェーンなどを応用した基盤技術に関する研究」というテーマで研究を行う。

非文字資料研究において研究者と一般の資料提供者が協力して資料の収集整理を行い、その研究成果を社会に発信し還元するためには、「資料の関連性や作業内容に即した検索とマイニング」、資料提供者や研究者の個人情報や重要情報、著作権の管理、資料提供や作業の対価やインセンティブとなる「多様な価値観に基づく地域通貨的価値交換」が必要となる。本研究では、「知識とサービス、物の流通と価値交換」、「知識とサービスの検索とマイニング」、「個人情報や重要情報、著作権の管理」で必要な基盤技術に機械学習とブロックチェーンなどを応用する。具体的にはアクセス制御に必要な資料間の関係性や電子透かしで必要な画像固有の情報の抽出に機械学習を利用したり、流通過程のコンテンツの作成、登録、利用、譲渡、二次利用などの時系列をとまなう事象の発生をブロックチェーンを利用して信頼できる第三者を仮定することなく行うことなどが挙げられる。

研究業績

- 1.KINOSHITA Hirotugu, MORIZUMI Tetsuya, “Access Control Model for the Inference Attacks with Access Histories”, Proc. of IEEE COMPSAC 2014,10.1109/COMPSAC.2017.41, Jul. 2017.
2. 宋 先波、森住哲也、木下宏揚（神奈川大），“テキストの相互情報量により非文字オントロジー間を接続する概念の提案とそのケーススタディ～潜在的テキストを確率変数として見る場合～”，信学技報、vol. 117, no. 126, SITE2017-18, pp. 135-140, 7月, (2017)。
3. 矢田一貴、木下宏揚、森住哲也，“トピックモデルを用いた潜在的関連性を考慮した情報検索”，第80回情報処理学会全国大会、7B-03、2018年3月。
4. 松下智樹、木下宏揚、森住哲也，“機械学習を用いる推論を考慮した情報漏洩の検出”，第80回情報処理学会全国大会、2E-06、2018年3月。
5. 森住哲也、木下宏揚，“確率的セキュリティモデルの可能性について”，2018年暗号と情報セキュリティシンポジウム (SCIS2018)、1C2-2、2018年1月。
6. 森住哲也，“論理学的存在者から見る確率的存在者の倫理とは何か”，信学技報、技術と社会・倫理研究会 (SITE)、3月、(2018)。
7. 辻順平、能登正人，“テーマパーク問題におけるパレート最適性を考慮した滞在時間短縮フレームワーク”，人工知能学会論文誌、Vol.33、No.2、pp.C-H98-1-9 (2018)。
- 8.Zhaoxiong Meng, Tetsuya Morizumi, Sumiko Miyata, Hirotugu Kinoshita, “Design Scheme of Copyright Management

- System Based on Digital Watermarking and Blockchain”, 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Vol.02, pp. 359 – 364 (2018)
9. 森住哲也、“論理学的存在者から見る確率的存在者の倫理とは何か”、信学技報、vol. 117, no. 471, SITE2017-76, pp. 213 – 219, 3月,(2018年)。
10. 森住哲也、木下宏揚、“確率測度空間に於いて脱構築装置を内在するアクセス制御について”、信学技報、vol. 118, no. 152, SITE2018-28, pp. 281 – 287, 2018年7月。
11. 中谷憲、森住哲也、木下宏揚、“ベイジアンモデルによる情報漏えい分析のための機械学習”、電子情報通信学会ソサイエティ大会、A-12-1、2018年9月。
12. 森住哲也、“確率測度的テキストと決定論的テキストを循環させる逆強化学習システムの倫理とは何か”、信学技報、vol.118。
- 13.no.345, SITE2018-62, 2018-12-06, pp.17 – 23, (2018)。
14. 森住哲也、“情報理論的尺度に基づく家族的類似クラスターの順序関係の学習可能性について：ベイジアン逆強化学習の報酬を家族的類似度と見做すこととは何か？”、2019-02-28 (SITE, IA), (2018)。
- 15.MENG Zhaoxiong, MORIZUMI Tetsuya, KINOSHITA Hirotsugu, MIYATA Sumiko,“Perceptual Hashing based on Machine Learning for Blockchain and Digital Watermarking”,IEEE 2019 Third World Conference on Smart Trends in Systems Security and Sustainability (WorldS4) DOI: 10.1109/WorldS4.2019.8903993 (30-31 July 2019)
16. 紅林宏祐、森住哲也、木下宏揚、“Boid 的アノテーションと Labeled-LDA による家族的類似の推論規則生成：推論攻撃分析と covert channel 攻撃分析を統合する機械学習のアプローチ”、信学技報、vol. 119, no. 141, SITE2019-36, pp. 243 – 249, 2019年7月。
17. 森住哲也、“必ずしも完全に分有されないロゴスと言語ゲームをつなぐ確率的存在者：セキュリティモデルの限界と人工知能の可能性”、信学技報、vol. 119, no. 141, SITE2019-43, pp. 317 – 324, 2019年7月。
18. 森住哲也、“論理空間の存在者及び確率測度空間の中の存在者として存在する事：身体性の必要条件としての言語ゲーム上の分有ロゴス”、信学技報、vol. 119, no. 329, SITE2019-86, pp. 41 – 46, 2019年12月。
19. 紅林宏祐、森住哲也、木下宏揚、“2つの確率モデルの組み合わせによる Multi-Label Learning の解釈について：ラベルを教師データとみなし、評価テキストを解釈するアプローチ”、信学技報、vol. 119, no. 329, SITE2019-81, pp. 7 – 12, 2019年12月。
- 20.Zhaoxiong Meng,Tetsuya Morizumi,Sumiko Miyata,Hirotsugu Kinoshita,“A Scheme of Digital Copyright Management System Based on Blockchain and Digital Watermarking: Research on Improvement Method of Perceptual Hashing based on Machine Learning”、信学技報、vol. 119, no. 329, SITE2019-83, pp. 21 – 27, 2019年12月。
21. 中谷 憲、孟 昭雄、森住哲也、木下宏揚、“畳み込みニューラルネットワークを用いた知覚ハッシュのための中間層の分析”、信学技報、vol. 119, no. 329, SITE2019-80, pp. 1 – 6, 2019年12月。

謝辞

芝浦工業大学の卒業研究において、非文字用のオントロジー構築とこれに基づく検索システムの構築に多大なる貢献をし、只見町の現地調査においても活躍した、生方優希さんに謝意を表する。