

第5班

非文字資料の効率的な検索と安全な流通

(1) 共同研究員名

研究班代表：木下宏揚

共同研究者：佐野賢治 能登正人 松澤和光 宮田純子 小松大介 鈴木一弘

(2) 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。

上記の目的達成に必要な基盤技術の提案を行い、その後、只見民具カードを対象に基本的なシステムを構築する。

最終的な成果目標は下記の通りである。

1. 非文字資料に特化したオントロジーを構築し精度の高い検索と新しい知見のマイニングを行うシステムを構築する。
2. 非文字資料のオントロジー構築、検索などに適したユーザインタフェースを構築する。
3. 非文字資料の検索、流通時に個人情報や機密情報を保護し、著作権の調停を自律的に行う流通システムを構築する。
4. 非文字資料の資料の作成、データ処理、資料の流通などを円滑に行うために地域通貨的決済手法を提案する。
5. 非文字資料とことば工学のコラボレーションおよび非文字資料からの会話文生成システムの提案を行う。

(3) 活動経過

2011年度

1. 非文字資料のオントロジーとマイニング

本研究資料は、民俗文化をベースとしていることから、同じものを指し示す場合でも、地域や年代によって相違が生じる。そのため、非文字資料の情報共有・情報流通には情報資源に関する情報、すなわち、メタデータを用いた意味情報検索が求められる。現在の Web では、Web 上に散在している情報資源を表現するメタデータを利用した高度な処理を行うことはできない。一方、Semantic Web

では、情報資源間の関係を構造的に表現したメタデータから、意味的な検索や推論・演算といった知的な処理を提供することができる。したがって意味情報検索を主体とする非文字資料の情報共有・情報流通に適すると考える。本研究では、民具資料カードを用いた「Ontologyを用いた民具のデータベース化」の研究で、非文字資料のOntologyを構築し、研究者に対し意義のある新たな知見の提示が可能であることを示した。しかし、具体的な推論機構については未解決であった。そこで、非文字資料のOntologyにJenaを用いたRDFの推論を導入することで、明示されていない関係を導出する。それにより、新たな関係を発見することができ、非文字資料のOntologyの有効性を実証した。

2. 非文字資料のためのユーザインタフェース

キュレーションという言葉がある。キュレーションとは、博物館や図書館の学芸員を「キュレーター (curator)」とも呼んでいるが、「情報を収集し、選別し、意味付けを与えて、それをみんなと共有すること」という定義である。つまり、情報の洪水の中で、それぞれのユーザにとって必要なものをコンテキスト (意味の文脈) に沿ってピックアップし、整理し、新しい付加価値を与えることである。クラウド内の大量にあるファイルの種類を自分なりにキュレーションし、「ファイルの見える化」を行い、ユーザにとって見やすいファイルシステムを提案した。

3. 多様な価値観を反映できる地域通貨的決済

非文字資料の研究者同士や資料の利用者間でサービスや資料を円滑に流通させるために、情報カプセルと価値ベクトルを用いた多様な価値観を反映できる地域通貨を提案した。

4. 非文字資料の著作権保護

近年、電子書籍の発達や、pixivなどの多型のイラスト投稿サイトの影響でインターネット上に多くのイラストコンテンツが存在している。それらの著作権を保護するために、不正なコピーを検出する手法が必要となっている。また、不正利用者はオリジナルのイラストをそのまま使うだけでなく、様々な処理によって改変を加える他、手書きによる複製や模写 (本論文ではトレースという言葉を用いる) したものを自分の作品として発表することが多い。これらの問題に対処するため、電子透かしなどの技術が存在するが、電子透かしは対象のコンテンツに著作権情報を埋め込む技術で、線画などの冗長性の少ないイラストに対しては人間に知覚できないように埋め込むことが困難となっている。線画の著作権保護の研究では、線画の著作権保護のための部分的複製検出法が提案されており、その有効性が証明されているが、カラーのオリジナルイラストから輪郭線部分を複製したものや、線画を元に着色を施した複製イラストに対する有効性は検証されていない。そこで本論文では、イラストから線画を抽出したものに対し、計算速度が早いSURF、線画像に対して有効であるHOGという局所特徴量を用いてデータベースを作成し、複製の疑いがある画像の特徴量と画像検索技術を用いて照合することで、著作権に違反している不正コピーを検出する手法を提案した。更に実験により、改変を加えた複製、トレース画像に対する提案手法の有効性を検証した。

2012 年度

5. 振る舞いに基づくアクセス制御と情報検索

新しい概念の情報漏えい防止と情報検索の要素技術となる“ふるまい”について考察を行った。家族的類似によって集まるふるまいの群れを前提とする倫理に関してウィトゲンシュタインの論理空間と言語ゲームでは〈私〉は消去される。本研究は〈私〉を映し出す多元言語ゲームの構造を定義し、倫理を「示す」ことを可能にする。アクセス制御行列上で、個々の〈私〉が〈私のふるまい〉を家族的類似の特性を使って集め、〈私〉が〈私のふるまい〉の境界を定義する。社会の言語的なふるまいは、ふるまいの集まりの群れによって記述される。そしてそのふるまいの群れは、群知能によって制御される。多元言語ゲームを制御する統合的な意味論は存在せず、自律分散的なメタヒューリスティクスが全体を調和させる。倫理とは、“公と私の価値循環が調和するように、個々の個人が制御と了解を全体に対して試みること”、と示されることである。

6. クラウドファイルシステム

「情報を収集し、選別し、意味付けを与えて、それをみんなと共有すること」という意味でのキュレーションを実現するために非文字資料のユーザインタフェースとして、群知能を用いたクラウドファイルシステムの表示方法を Android 端末に実装を行った。クラウドファイルシステムでは、注目しているコンテキストに応じて、重要なファイル・関連があるファイルほど互いに集まり、群れを作るようにする。また、ファイルや情報の、「人間の扱い方『ふるまい』」に注目する。このようなふるまいが集まると、他者との連携により人間の創発的活動（学術的活動・ビジネス・地域ボランティアなど）を刺激・支援するシステムを作ることも可能になる。

7. 自由曲線を用いた電子透かし

著作権保護に関しては、自由曲線を用いた電子透かしの提案を行った。現在パソコン、およびインターネットの普及、小型端末の登場により、ネットからダウンロードして、パソコンや携帯端末などで読むことの出来る電子書籍の市場が広がり始めている。小説などの文章だけでなく、漫画コンテンツなども電子書籍化が始まっている。また、ネット上には様々な種類の画像投稿サイトも存在しており、アマチュアの絵描きの自由な投稿も行われている。それに付け加え、漫画を書く際にアナログではなく、パソコン上で原稿を作成する漫画家も増え始めた。非文字資料研究においても「日本近世生活絵引」など二値の線画像で構成されている資料が多数存在している。しかし、デジタルデータは複写、保存、加工が容易に行えるという特徴をもっており、ペイントソフトを使えばトレースなども容易に行うことが出来る。さらにはネットワーク技術の発展により、これらのデータを複数のユーザに対して配信することが容易になった。そのために、デジタルデータのコンテンツが意思に反して利用されることが起こりうる。このような問題を解決するために電子透かしなどの著作権保護技術が用いられている。しかし、漫画など主に二値の線により構成されている画像に対しては、画素数に対する情報量が少ないので品質を保ったまま必要な情報を埋め込むことは困難である。本研究では、二値の線画の著作権管理に関して、アクセス制御としての DRM や、デジタル・フォレンジックとしての特

微量データベース、電子透かしについて、有効性の検討を行う。さらに、それらを踏まえ二値の線画に適したベジェ曲線を用いた電子透かし構成法について提案した。

8. 地域通貨的決済手法に群知能を用いて最適な決済を行う手法

地域通貨的決済手法に群知能を用いて最適な決済を行う手法を提案した。研究資料などの情報リソースや資料分類作成などのサービスを円滑に流通させるために、多様な価値観を反映可能な地域通貨的価値交換システムを提案してきた。二者間の取引の際に取引後の両者の価値ができるだけ多くなるように、取引に用いる価値の最適な組み合わせを求めることは困難な問題である。本研究では粒子群知能最適化手法（PSO）を用いて、決済に用いる価値ベクトルの最適な組み合わせを求める手法を提案した。

9. 推論による情報漏えい防止のためのハイパーグラフモデル

推論による情報漏えい防止のためのハイパーグラフモデルの提案を行った。インターネット技術等の発展に伴い、人類が創出する情報量は爆発的に増加している。また、SNS やクラウドサービスの台頭により、それらの情報へのアクセスおよび解析が容易になった。特に近年は、ライフログと呼ばれる個々人のデジタル活動記録の有効活用に注目が集まっている。このような情報爆発時代においては、蓄積された膨大な情報群を解析することによって秘密情報を得ようとする攻撃に対応できる新しい情報漏えい対策やプライバシー保護対策が必要となる。従来は、個人情報や企業機密のような秘密情報がそのまま丸ごと漏えいしないようにすることであった。しかし、一つ一つの情報それ自体は秘密情報ではなくても、それらが複数集まり何らかの推論を施すことによって、秘密情報を抽出できてしまうこともある。そこで本研究では、ハイパーグラフを用いてこの問題をモデル化することを提案し、推論攻撃に対する安全性を定義したうえで、提案モデル上で考えるべき問題について検討した。

2013 年度

10. 只見町インターネット・エコミュージアムのキーワード検索の改善

専門的な分野について情報検索する場合、専門的な知識を持たなければ関連性が深い項目を見逃すなど有効な情報検索が行えない。只見町インターネット・エコミュージアムでは、民具の公開を行っている。しかしながら、そこにある検索機能はユーザの意図した検索キーから正確な情報を出力しないという問題点がある。そこで、非文字資料の Ontology にウェブ・オントロジー言語 OWL を用いることで、明示されていない関係を導出する。それにより、新たな関係を発見することができ、本当に必要な情報を的確に検索することが可能となる。

11. 自己組織化可能な群知能を用いた情報リソースの管理

ビックデータやクラウドで知られている様々な情報データは、ファイルの持つ複数の属性により、管理が複雑になる。また、現在は木構造がファイル管理の主流であるが、直感的にファイルの位置を

把握することが難しい。一方で、多数のファイルであっても、類似性の高いファイルごとに大きく分類されていれば、直観的にユーザがデータを扱えると考えられる。そこで本研究では、各データを、自己組織化可能な群知能データとみなし、それらの群知能データを制御することで、視覚的にデータを管理可能にする。提案管理手法は、データ管理の過程に応じて自己組織化する手法にも適用可能である。

12. ACO を用いた検索過程を重視した検索手法

インターネットで検索をするときに、比較的簡単に一定レベルの検索結果は得ることができる。しかし、検索の質は常に十分とは限らない。もしその情報について更に深く理解しようとするれば、関連した情報についても調査する必要がある。そのとき、その分野に詳しい人の検索結果やその過程を参考にできれば、より詳細で多くの情報を検索過程から得ることができる。本稿では、検索過程の質を考慮した検索手法を提案する。群知能の ACO を用いた手法と「推薦システム」と呼ばれるシステムを用いた手法の二つを提案する。

(4) 研究成果

主な学会発表

1. 佐藤俊輔、木下宏揚、松澤和光：オントロジーを用いた民俗学研究者支援システム、FIT2011 第 10 回情報科学技術フォーラム、4E-2、2011 年 9 月 8 日発表。
2. KINOSHITA Hirotsugu, TAJIMA Yoshiaki, KUBO Naoya, MORIZUMI Tetsuya, SUZUKI Kazuhiro: "A local currency system reflecting variety of values" International Symposium on Applications and the Internet (IEEE/IPSJ), ITeS 2011, pp. 562-567 (2011-7).
3. 工藤護、工藤敬、木下宏揚、森住哲也：“イラストの著作権保護のための局所特微量を用いた複製検出”、暗号と情報セキュリティシンポジウム、4F1-4 (2011-2)。
4. 鈴木遼、鈴木一弘、森住哲也、木下宏揚：推論による情報漏えい防止のためのハイパーグラフモデル。電子情報通信学会論文誌 D, Vol. J95-D, No. 4 pp. 812-824 (2012-04-01)。
5. KINOSHITA Hirotsugu, TAJIMA Yoshiaki, MORIZUMI Tetsuya, NOTO Masato, KAMINABE Hideyuki: "A local currency system reflecting variety of values with a swarm intelligence", Applications and the Internet (SAINT), 2012 IEEE/IPSJ 12th International Symposium, 16-20 July. 2012.
6. 工藤敬文、宮田純子、森住哲也、木下宏揚：自由曲線を用いた電子透かしの提案、信学技報、vol. 112, no. 343, SITE2012-44, pp. 7-11, 2012 年 12 月。
7. 田島佳明、宮田純子、森住哲也、木下宏揚：地域通貨的価値を利用した価値の交換システム、信学技報、vol. 112, no. 343, SITE2012-43, pp. 1-6, 2012 年 12 月。
8. 久保直也、森住哲也、鈴木一弘、能登正人、木下宏揚：“群知能を適用したアクセス制御システム”、暗号と情報セキュリティシンポジウム scsi2012, (2012-01)。
9. 森住哲也、久保直也、鈴木一弘、能登正人、木下宏揚：“群知能を言語ゲームと見做すアク

- セス制御”、暗号と情報セキュリティシンポジウム scsi2012, (2012-01)。
10. Hirotsugu Kinoshita, Yoshiaki Tajima, Tetsuya Morizumi, Masato Noto, Hideyuki Kannabe and Sumiko Miyata: A Local Currency System Reflecting Variety of Values with a Swarm Intelligence, Proc. of The 12th IEEE/IPSJ International Symposium on Applications and the Internet, pp. 251-255 (2012).
 11. Hideyuki Kannabe, Masato Noto, Tetsuya Morizumi and Hirotsugu Kinoshita: Agent-Based Social Simulation Model that Accommodates Diversity of Human Values, Proc. of 2012 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1818-1823 (2012).
 12. 中村峻生、鈴木一弘、森住哲也、宮田純子、木下宏揚：推移閉包アルゴリズムを用いた covert channel 検出”、電子情報通信学会論文誌 D、Vol. J96-A, No. 4, pp. 175-183, Apr. 2013。
 13. Sumiko Miyata, Hirotsugu Kinoshita, Tetsuya Morizumi, “A model of the value exchange system in the university with the big boss game,” In Proc. of ICEC 2013, Aug. 2013.
 14. Sumiko Miyata, Hirotsugu Kinoshita, Tetsuya Morizumi, Li Chao, “Game theoretic analysis of the value exchange system,” In Proc. of IEEE International Symposium on Applications and the Internet, Jul. 2013.
 15. 小泉駿、石田克憲、宮田純子、森住哲也、木下宏揚、“ACO を用いた検索過程を重視した検索手法”、信学技報 SITE2013, SITE2013-57, vol. 113, no. 355, pp. 163-168, 2013 年 12 月。
 16. 石田克憲、小泉駿、宮田純子、森住哲也、木下宏揚、“自己組織化可能な群知能を用いた情報リソースの管理”、信学技報 SITE2013, SITE2013-56, vol. 113, no. 355, pp. 163-168, 2013 年 12 月。

(5) 今後の課題と展望

目的 1 については、非文字資料のオントロジー構築方法と、これに基づくデータの間の関係の推論手法を提案し当初の目的を達成できた。目的 2 については、ネットワーク上に配置されている様々な情報リソースを、多角的な視点から検索可能なクラウドファイルシステムとそのインタフェースを提案し、目的を達成できた。目的 3 については、個人情報保護については、人の行動などを表す「ふるまい」に基づく新しいアクセス制御法を構築し利便性と安全性を両立可能とした。また、イラストなどを対象とした電子透かし構成法を提案し、当初の目的は達成できた。しかし、著作権の自律的な調停については、基礎的検討を行ったのみで具体的な提案には至らなかったため今後の課題となる。目的 4 については、地域通貨的決済手法を提案し、ゲーム理論を用いてその有効性を確認できたので当初の目的は達成できた。目的 5 については、構想段階にとどまり、提案には至らなかったため今後の課題としたい。

今後の課題と展望としては、第二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、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族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あげられる。具体的には以下の項目からなる。(1) オントロジーを用いたデータマイニングの実際

の資料に対する応用、(2) 資料のデータベースをクラウド化する際の個人情報保護と著作権管理、(3) 資料の整理とデータ入力や流通を円滑に行うためのゲーム理論や群知能などに基づく価値交換モデルの構築、(4) 資料を直感的に取り扱うことを可能とする、群知能を用いた操作のコンテキストに基づくユーザインタフェースの構築、以上の4点である。