

令和 4 年 6 月 20 日現在

機関番号：10101

研究種目：基盤研究(B)（一般）

研究期間：2018～2021

課題番号：18H03338

研究課題名（和文）多様なニュースサイトの比較による報道スタンスの分析と可視化の研究

研究課題名（英文）Research on Analysis and Visualization of News Site Stances based on the Comparison of Multiple News Sites

研究代表者

吉岡 真治（Yoshioka, Masaharu）

北海道大学・情報科学研究院・教授

研究者番号：40290879

交付決定額（研究期間全体）：（直接経費） 13,100,000円

研究成果の概要（和文）：本研究では、ニュースサイトのスタンスを分析する方法として、トピックごとの記事の賛否に注目して多面的な分析を行う方法を提案した。具体的には、ニュースサイトには、右翼・左翼のような一元的なスタンスだけではなく、ニュースジャンルやトピックに関する興味の違いがあることを確認したことを受け、これらの興味の違いを考慮したニュースサイトのスタンスの類似性を利用者の興味を考慮しながら分析できるインタラクティブなシステムの提案を行った。また、並行して行った固有名詞に関する知識として、Wikipediaカテゴリの情報を知識工学的な観点から整理したオントロジーの提案をおこなった。

研究成果の学術的意義や社会的意義

本研究では、これまでの研究やニュースサイトの分析で行われてきた右翼・左翼といった一次元的なスタンスではなく、トピックの興味を考慮して多面的にスタンスを分析するという考え方に基づいて、情報科学の観点からニュースサイトのスタンスを分析する新しい方法を提案したという点に学術的な意義がある。現時点では、その解釈が難しい状況ではあるが、現在問題となっている偏向報道などについての分析を行うための基盤となりうる技術である社会的意義は大きいと考えている。また、並行して開発したWikipediaカテゴリオントロジーは、知識工学的な観点からWikipediaの情報を整理したものであり、今後の活用が期待される。

研究成果の概要（英文）：In this research, we proposed a method to analyze news site stance of particular news topic using the distribution of polarity of the articles published by the news site from multiple perspectives. In this research, we confirmed that it is important to understand the interests on subtopics and/or genres of the news articles for detailed analysis of the news site stance. Based on this understanding we proposed a system that can analyze news site stance interactively by taking account the polarity of the related news articles for multiple subtopics and/or genres.

We also proposed Wikipedia category ontology that is new knowledge resource extracted from Wikipedia category from the viewpoint of knowledge engineering. This new knowledge can be used for analyzing the type of the named entity.

研究分野：知識工学

キーワード：テキストマイニング 報道分析 可視化 オピニオンマイニング オントロジー

## 1. 研究開始当初の背景

研究の着想を得た2010年代には、ニュース記事へのアクセスが、これまでの新聞やテレビ放送などを用いた購読・視聴が、インターネットを介した購読・視聴へと大きく切り替わっていく時代であった。また、そのインターネットを介したアクセス手法についても、信頼性の高い特定のニュースサイトの記事を中心に購読するスタイルから、Yahoo!ニュース、Google News に代表されるニュースアグリゲーションサイト、Smartnews や Gnosy のようなニュースアグリゲーションアプリを用いた複数のニュースサイトからの記事の中から機械的なアルゴリズムで推薦されたものを読むというスタイルへと変化してきていた。さらに、Twitter、Facebook などの SNS からの記事の推薦などが盛んになると、公平で信頼性の高い記事よりは、話題になりやすい記事を配信するようなニュースサイトが増えてきていた。

このような問題に対して、人手によりニュースサイトの特性を分析する活動として、Media bias chart (<https://adfontesmedia.com/>)などの活動が開始されていた。しかし、多数のニュースサイトを人手で分析することは困難であり、新しく立ち上がるニュースサイトなどを継続的に分析していくためには、計算機を用いた分析を行う枠組が望まれると考えた。

また、研究代表者の吉岡と共同研究者の神門は、これまでに、「異なる特徴を持つニュースサイトを比較対照する世界ニュース分析システムの研究」などを行った実績があり、このような比較対照により、ニュースサイトの特性を明らかにできるのではないかと考えた。また、研究代表者がマサチューセッツ大学アマースト校の Allan 教授（研究協力者として本研究に参加）を訪問した際に、Twitterなどを対象とした「物議を醸す(Controversial)話題」についての研究について話を伺い、特定の話題に対する賛否を分析することで、ニュースサイトのスタンスについての分析を行うことができるのではないかと考え、本研究計画を策定した。

## 2. 研究の目的

本研究では、様々な事象について、世界中に存在する多様なニュースサイトを比較分析することにより、各ニュースサイトについて賛否や興味のあるトピックに基づいた定量的な報道スタンスを明確化する方法を提案するとともに、ニュースサイト間の関係を可視化することで、俯瞰的に事象を分析するシステムの提案を行う。具体的には、世界各地の様々なニュースサイトから収集したニュース記事について関連する人物やトピック、賛否の度合いなどの情報を付加して構築されている GDELT project のデータベースを活用し、ニュースサイト毎の賛否の度合いに関する情報の計算に利用する。また、ニュースサイト毎に興味を持つサブトピックの違いやそれに対する賛否の違いなどを分析することにより、報道スタンスを定量的に表現する方法を提案する。また、俯瞰的な分析を行うために、これらの基準を用いてニュースサイトの類似性(非類似性)の尺度を用いて可視化すると共に、利用者の興味に応じた可視化を行う方法についても提案する。

## 3. 研究の方法

一般に、世界中のニュース記事を大量に入手し解析することは、非常に困難であるが、世界中の新聞記事を収集分析し、その動向を分析するためのデータを構築するプロジェクトである GDELT project によるデータ公開が行われている。GDELT project では、Google と連携する形で大量のニュース記事を分析し、その記事の発信者やその URL、記事中に含まれている名前や組織といった固有名詞などの情報、賛否に関するトーンなどの情報を付与したデータベースとして公開している。本データベースは、著作権の関係などから本文の情報を得ることはできないが、大量の記事の中から、たとえば、Trump 大統領についての言及を含む記事だけを選択し、発信者ごとの賛否について分析することが可能である。本研究ではこのデータベースに収録されているニュースサイトを対象として分析を行うシステムを提案することとした。

また、提案システムにおいて、上記の研究目的を実現するために、次のような項目について順番に検討を行うことで研究を進めた。

- ① 論文の賛否についての分析
- ② 論文の賛否の割合に注目したニュースサイトのスタンスの分析
- ③ トピックに注目したニュース記事の分類とニュースサイトの特徴の表現
- ④ ユーザの興味に応じた類似性を考慮したニュースサイトの可視化手法の提案

また、この結果を踏まえて、スタンスの分析に適切なトピックの決め方や、論文の賛否に関するより詳細な分析を通して、システムの改良を目指した。

さらに、上記に並行する形で、新聞記事などに出てくる固有名詞などの分類に有用となる知識を Wikipedia の記事分類に用いられている Wikipedia カテゴリを利用する方法と、ニュース記事の意見を詳細に分析する際に有効となると考えられる言説の含意関係の分析についても、並行して研究を行うこととした。

#### 4. 研究成果

研究の第一段階では、2016年の米国の大統領選挙に関する記事として、二人の候補者 Donald Trump と Hillary Clinton についての言及のある記事を用いて賛否の分析を行うこととした。初期の想定としては、保守派（右翼）の支持層に対応するニュースサイトでは、Trump 候補に賛成が多く、リベラル（左翼）の支持層に対応するニュースサイトでは、Clinton 候補の賛成が多いといったシンプルな対応関係になるのではないかと想定したが、新聞記事の平均的なトーン（賛否）が否定的な記事が多い影響なども受けて、賛成の記事の絶対数や単純な割合では、適切な分析ができないことが判明した。

この問題に対し、平均的な賛否の割合というのが想定できるのであれば、Trump 候補に関する記事と Clinton 候補に関する賛否の割合の違いを比較することにより、サイトの特徴が分析できると考えた。具体的には、図1に示すように、Trump, Clinton 両候補に関する記事（両者について言及している記事は、どちらに対する意見かが分からないため分析対象から除外）を対象に、賛否の割合を可視化した。図1の左は Trump 候補、中央は Clinton 候補に関する分析記事であり、縦軸と横軸は、それぞれ否定的な記事と肯定的な記事の割合を表している。この可視化結果から明らかなように、全般的に否定的な記事の割合が多い（左上に点が多い）ことがわかる。この中でも、たとえば、青い丸で表示されている [freerepublic.com](http://freerepublic.com) は Trump 候補については、青い線が Clinton 候補の場合に比べて右側にあることから、Trump 候補に対して、Clinton 候補を支持していると考えられる。図1の右は、この違いを可視化するために、縦軸と横軸に、Clinton 候補と Trump 候補それぞれに対する肯定的な記事の割合と否定的な記事の割合の差を表している。各々の候補について、同じような割合であれば、真ん中の斜めの直線に近いところにサイトが表示され、直線から左上に離れるほど、Clinton 候補の支持バイアスが強く、逆に右下に離れるほど Trump 候補の支持バイアスが強いと判断できる。先ほどの [freerepublic.com](http://freerepublic.com) は、この表示においても、Trump 候補の支持バイアスが強いと判断できる。さらに、この可視化において、真ん中の直線から遠いところを見ると、Clinton 候補に対しては、[sci-tech-today.com](http://sci-tech-today.com) のような極端に Trump に否定的な記事が多いサイトや、[thebostonpilot.com](http://thebostonpilot.com) といったより極端なサイトを発見できることが確認された。

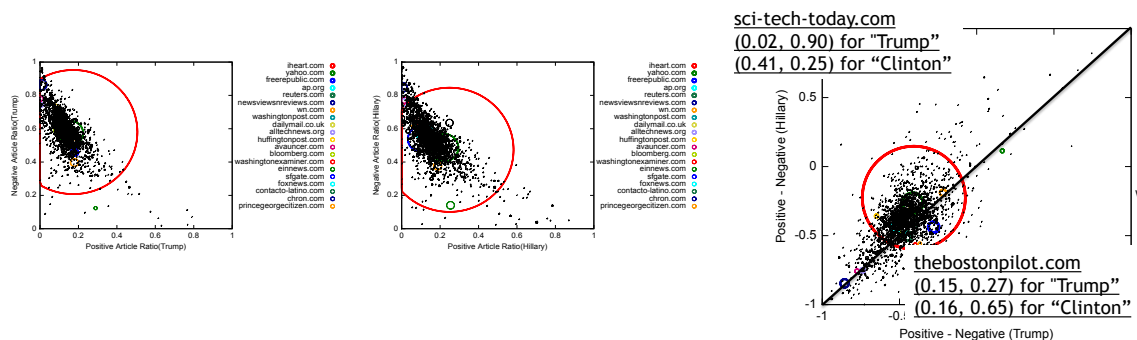


図1：Trump の記事と Clinton の記事の各サイトの賛否の割合と分布の比較のための可視化

この結果、GDELT project の賛否の情報を用いたニュースサイトの分析は、特徴的なニュースサイトを見つけることに有効であると判断した。一方、分析の対象としたサイトの中には、国内の問題にしか興味がないサイトや、経済、外交といった話題に特化したようなサイト、広く多くの話題について扱う総合的なニュースサイトなど様々なニュースサイトが存在し、賛否について似たような傾向を持っていても、話題に対する興味が異なるようなサイトが多く存在することが確認された。

このようなニュースサイトごとの興味を考慮したスタンスの分析を行うために、一般的なニュースサイトでよく用いられているジャンル分類を考慮した分析を行う枠組みを考えた。ただし、今回対象とする GDELT project で公開されているデータについては、本文の情報が存在しないという問題点があり、この問題を解決するために、記事の URL を利用する方法を提案した。多くのジャンルを設定しているニュースサイトにおいては、URL として <https://サイト名/ジャンル/記事 ID> のような形式を利用していることが多く、サイトごとのジャンルの読み替え規則を作ることでジャンルの推定が可能になる場合がある。このようにジャンルが特定されないニュースサイトにおいても、英文のニュースサイトでは、記事 ID が本文のタイトルをもとに、空白を「-」で置き換えたような名前を利用することが多く、タイトルに関する情報が得られることが多いことが確認された。本研究では、このような URL から用いられる情報を用いて、ニュース記事を「政治」・「国内」・「国際」・「経済」・「科学」・「ライフ」・「スポーツ」・「エンターテインメント」の7種類に分類することで、ニュースサイトごとの興味を表現することとした。

具体的には、URL にジャンル情報を含むサイトについては、読み替え規則を利用することで、各記事を上記の7種類から対応するジャンルを決定し、タイトルに関する情報を含む場合には、深層学習に基づく機械学習が可能な BERT を用いて分類した。この手法では、ジャンル情報とタイトルの両者を含む URL から作られるジャンルと本文の対応関係を訓練データとして用いる

ことで、タイトルを与えた場合にジャンルを出力するシステムを構築した。本研究では、Donald Trump に関する記事に限定したことも影響したとは考えているが、87%の精度で適切な分類を推定できるシステムが構築された。

本研究では、上記の分類で分類可能な URL の情報を持ち、Media bias chart で、右翼・左翼の情報を確認することのできる 77 のニュースサイトを対象として、ジャンル分類を行い、ニュースサイトのスタンス分析を行うこととした。本枠組みでは、従来のニューススタンスの分析が肯定的な記事と否定的な記事の割合にのみ注目していたのに対し、賛否の 3 分類（肯定・中立・否定）とジャンルの 7 分類の組み合わせである 21 次元のベクトルで各ニュースサイトの特徴を表現し、このベクトルの類似度でニュースサイトの類似度を測ることとした。また、21 次元のベクトルのままでは、可視化が難しいため、多次元尺度構成法（MDS）を用いた次元圧縮を行い、類似度が高いニュースサイトを近くに配置するような可視化方法を考えた。

また、一般的な MDS の結果では、ある種の情報量の基準に基づく画一的な結果を表現することになるが、これは、必ずしも、利用者が想定している類似度と対応しない場合が考えられる。たとえば、経済に興味のある利用者は、経済のジャンルの記事に関する情報を用いた類似度に重きをおいた表現をすることが望ましいと考えた。

そのため、本研究では、距離尺度学習の枠組と MDS の可視化をインタラクティブに行うことで、利用者の意図を推定しながら、利用者の意図に応じた可視化を実現できる方法である SIRIUS: Interactive Visualization Tool for Multidimensional Data (M. Dowling et. al., 2019) を用いた可視化システムを構築した。

本システムでは、利用者が考える類似した（あるいは、非類似の）ニュースサイト群を指定することで、利用者が興味のある特徴（各々の特徴量に対する重み）を推定するとともに、その特徴を考慮した類似度に基づく可視化を行うことで、利用者の興味を反映した類似度に基づくニュースサイト群の可視化が行える。また、本システムでは、ニュースサイトの類似度から操作を開始するだけでなく、興味のある特徴を直接、数値的にコントロールすることで、ニュースサイトの可視化を行うことも可能である。図 2 は、SIRIUS による可視化結果で、サイトの一般的なスタンスを示すために、Media bias chart における右翼（赤）・左翼（青）の分類を用いて表示している。

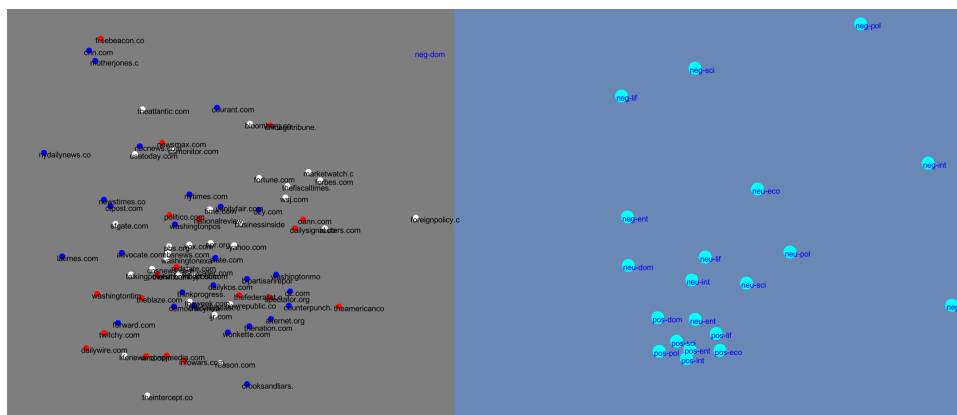


図 2：SIRIUS によるニュースサイトの類似度の可視化（左）と属性の類似度の可視化（右）

この結果から、右翼、左翼のサイトのいくつかをまとめて類似サイト・非類似サイトとして指定した結果をもとに、興味のある特徴を推定して可視化した結果を図 3 に示す。左側の図が右翼・左翼の一部のサイトを左上・右下に設定した後で、再度、ニュースサイト・属性を再配置した結果が図 3 の右の図である。

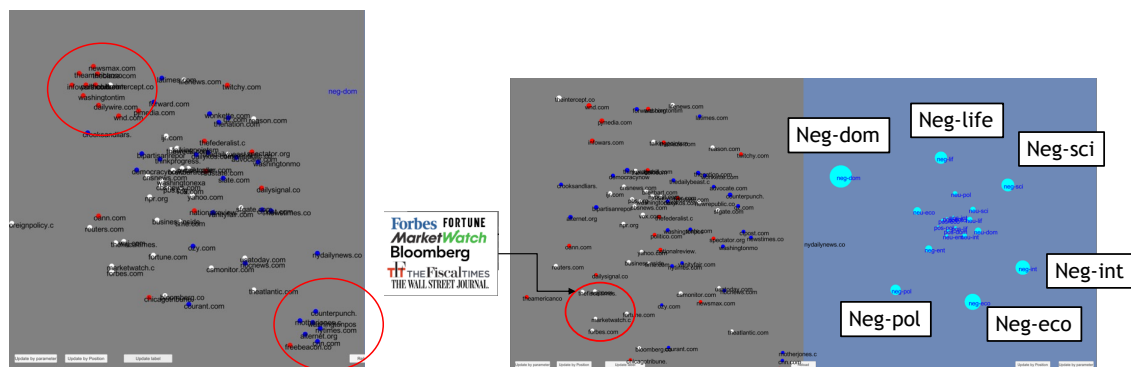


図 3：ニュースサイトの類似度の設定（左）とニュースサイト・属性の再配置（右）

特徴的な属性としては、各ジャンルにおけるニュース記事のうち否定的な記事の割合に重きを置く（大きな丸になっている）という結果になり、金融の記事に注目した中立的なニュースサイトがまとるといった現象も確認された。

また、別の可視化方法として、外交に興味があるといった形で、外交に関する肯定・中立・否定の重みを直接的に変更すると、外交に興味のあるサイト以外は、中央につぶれた形で表現され、外交に興味のあるサイトを中心とした可視化結果が表示された。

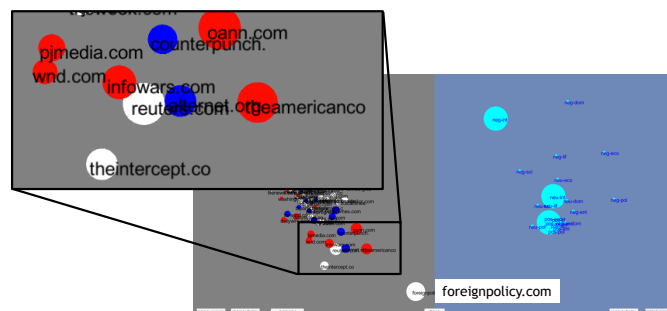


図4：外交に関する特徴に重きを置いた可視化結果

本手法により、ジャンルを考慮したインタラクティブな可視化を行うことができることが確認された一方で、詳細に記事を分析する際には、ジャンルを中心とした分析では、不十分ということが明らかになった。具体的には、多くの記事において、7つのジャンル分類は、それぞれ排他的な分類ではなく、国内の経済の話題といった分類を考えた方が良いという指摘があった。

この問題を解決したより良い分析を行うためには、ジャンルのような汎用的な分類ではなく、具体的な話題に応じた分析を行うことが望ましいと考えられた。

この問題に対応するために、実際にニュース記事の本文を活用したトピックの生成とそのトピックに応じた賛否の分析を検討した。具体的には、大量の記事から、分散表現としての **Topic Model** を作成した上で、クラスタリングを行うことにより、話題に対応する記事群と特徴語を抽出する枠組である **Top2vec: Distributed representations of topics (Angelov, 2020)** を用いた話題の作成と分析に関する予備実験を行なった。ニュース記事の本文については、**GDELT Project** のデータから事前に設定したキーワードを含む記事についてのみ、**URL** にアクセスして記事を取得した。

この実験の結果得られた話題は、その大きさをうまく制御しない時には、話題が比較的小さいサイズで作られるため、肯定（あるいは否定）の話題でよく用いられる単語などを考慮すると、ニュースサイトの特徴と関係なく肯定的（否定的）な記事の多い話題などが作成されてしまい、適切な分析には至らなかった。今後は、このような小さく分かれてしまう話題を適切に組み上げて、特定の話題についての肯定的な記事と否定的な記事を含む「物議を醸す(Controversial)話題」となる大きさの話題を作成する方法などを検討する必要があることが確認された。今後は、適切な話題が設定可能となれば、従来のジャンルの代わりとして、話題を用いることで、話題ごとのスタンスを考慮したより詳細な多角的なスタンス分析が可能になると考えている。

また、並行して行なってきた新聞記事などに出てくる固有名詞などの分類に有用となる知識を Wikipedia の記事分類に用いられている Wikipedia カテゴリを利用する方法の研究においては、具体的に日本語 Wikipedia のカテゴリ階層を分析し、カテゴリのタイプ（分類をあらわす Set カテゴリ、話題を表す Topic カテゴリ、また、それらの組み合わせ）とカテゴリ間の関係（推移律が成り立つクラス・サブクラス階層や、推移率の成り立たない制約としての利用など）を整理した基本語彙を提案するとともに、実際の 2017 年 10 月時点の日本語版 Wikipedia に含まれる Wikipedia の管理用カテゴリを除く全ての Wikipedia カテゴリとその関係について分類を行った Wikipedia カテゴリオントロジーを Linked Open Data の資源として公開した (<http://wcontology.org/>)。こちらの資源については、Linked Open Data Challenge 2021 でデータ作成部門 優秀賞とオントロジー賞（オントロノミー合同会社）を受賞した。

もう一件のニュース記事の意見を詳細に分析する際に有効となると考えられる言説の含意関係の分析の研究については、日本語の法律条文を対象とした言説の含意関係分析の課題 (COLIEE: Competition on Legal Information Extraction and Entailment) に取り組んだ。この課題では、司法試験の問題文 (択一問題) の言説が法律条文 (今回は民法を利用) により含意されるか (正しい言説か否か) を判断する課題である。このような含意関係認識においては、十分な訓練データが存在すれば深層学習をもちいたシステムの性能が向上することは知られているが、訓練データの作成にはコストがかかり、訓練データの拡張が必要とされている。本課題では、法律条文により言説が含意されるか否かを議論する際の観点が、条文で想定されている条件を問題の言説が満たしているかどうかを判断する意味的な判定と、最終的な法的な判断が一致しているか否かという論理的な判定という二つの観点があることに注目した。前者の意味的な判定については、どのような負例を作るのが適切かといった複雑な議論が必要であるのに対し、後者の論理的な判定については、法的な判断を反転させることで、機械的に負例を作成することが可能である。法律条文をそのまま用いるような正例に対して、この負例を組み合わせることで、大量の訓練データを作成することができる。このように拡張した訓練データを用いて学習を行なったシステムで含意関係認識を行ったところ、正解率が 70%と、参加者中、最も良い成績を示すシステムを構築できた。

## 5. 主な発表論文等

〔雑誌論文〕 計21件（うち査読付論文 15件 / うち国際共著 4件 / うちオープンアクセス 10件）

1. 著者名 Rabelo Juliano, Goebel Randy, Kim Mi-Young, Kano Yoshinobu, Yoshioka Masaharu, Satoh Ken	4. 巻 16
2. 論文標題 Overview and Discussion of the Competition on Legal Information Extraction/Entailment (COLIEE) 2021	5. 発行年 2022年
3. 雑誌名 The Review of Socionetwork Strategies	6. 最初と最後の頁 111-133
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s12626-022-00105-z	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Aoki Yasuhiro, Yoshioka Masaharu, Suzuki Youta	4. 巻 16
2. 論文標題 Data-Augmentation Method for BERT-based Legal Textual Entailment Systems in COLIEE Statute Law Task	5. 発行年 2022年
3. 雑誌名 The Review of Socionetwork Strategies	6. 最初と最後の頁 175-196
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s12626-022-00104-0	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Masaharu Yoshioka, Norihiko Tatsunami, Masahiko Itoh, Noriko Kando, and James Allan	4. 巻 -
2. 論文標題 Interactive Visualization of the Polarity-based Stance of News Websites using News Genres	5. 発行年 2021年
3. 雑誌名 Proceedings of the 9th International Workshop on News Recommendation and Analytics (INRA 2021)	6. 最初と最後の頁 62-69
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Masaharu Yoshioka, Youta Suzuki, and Yasuhiro Aoki	4. 巻 -
2. 論文標題 BERT-based Ensemble Methods for Information Retrieval and Legal Textual Entailment in COLIEE Statute Law Task	5. 発行年 2021年
3. 雑誌名 Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment (COLIEE 2021)	6. 最初と最後の頁 78-83
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Tetsuya Ishida, Yohei Seki, Wakako Kashino, and Noriko Kando	4. 巻 29
2. 論文標題 Extracting Citizen Feedback from Social Media by Appraisal Opinion Type Viewpoint.	5. 発行年 2022年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Masaharu Yoshioka and Takanori Nakagawa	4. 巻 -
2. 論文標題 Wikipedia Category Ontology: A Framework for Utilization of the Wikipedia Category Structure by Knowledge Engineers	5. 発行年 2020年
3. 雑誌名 Proceedings of the ISWC 2020 Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 19th International Semantic Web Conference (ISWC 2020)	6. 最初と最後の頁 300-305
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Rabelo Juliano, Kim Mi-Young, Goebel Randy, Yoshioka Masaharu, Kano Yoshinobu, Satoh Ken	4. 巻 LNCS12331
2. 論文標題 A Summary of the COLIEE 2019 Competition	5. 発行年 2020年
3. 雑誌名 New Frontiers in Artificial Intelligence. JSAI-isAI 2019. Lecture Notes in Computer Science,	6. 最初と最後の頁 34 ~ 49
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-58790-1_3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Teruko Mitamura, Masaharu Yoshioka, Tomoyosi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Kenichi Yokote, Tatsunori Mori, Kenji Araki, Satoshi Sekine and Noriko Kando	4. 巻 -
2. 論文標題 Overview of the NTCIR-15 QA Lab-PoliInfo-2 Task	5. 発行年 2020年
3. 雑誌名 Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies,	6. 最初と最後の頁 101-112
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Masaharu Yoshioka and Yoshiaki Koitabash	4. 巻 -
2. 論文標題 HUKB at SHINRA2020-ML task	5. 発行年 2020年
3. 雑誌名 Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies,	6. 最初と最後の頁 211-215
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Masaharu Yoshioka and Youta Suzuki	4. 巻 -
2. 論文標題 HUKB at COLIEE2020 Information Retrieval Task	5. 発行年 2020年
3. 雑誌名 Proceedings of the International Workshop on Juris-Informatics 2020 (JURISIN 2020) in association with the 12th JSAI International Symposia on AI (JSAI-isaI 2020)	6. 最初と最後の頁 288-297
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Keiichi Takamaru, Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Kotaro Sakamoto, Madoka Ishioroshi, Teruko Mitamura, Noriko Kando	4. 巻 -
2. 論文標題 Extraction of the Argument Structure of Tokyo Metropolitan Assembly Minutes: Segmentation of Question-and-Answer Sets	5. 発行年 2020年
3. 雑誌名 Proceedings of the 12th Language Resources and Evaluation Conference	6. 最初と最後の頁 2064-2068
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hiromu Onogi, Kiichi Kondo, Younghun Lim, Xinnan Shen, Madoka Ishioroshi, Hideyuki Shibuki, Tatsunori Mori, Noriko Kando	4. 巻 -
2. 論文標題 Forst: A Challenge to the NTCIR-15 QA Lab-PoliInfo-2 Task	5. 発行年 2020年
3. 雑誌名 Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies,	6. 最初と最後の頁 160-167
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -



1. 著者名 Kano Yoshinobu, Kim Mi-Young, Yoshioka Masaharu, Lu Yao, Rabelo Juliano, Kiyota Naoki, Goebel Randy, Satoh Ken	4. 巻 11717
2. 論文標題 COLIEE-2018: Evaluation of the Competition on Legal Information Extraction and Entailment	5. 発行年 2019年
3. 雑誌名 New Frontiers in Artificial Intelligence. JSAL-isAI 2018. Lecture Notes in Computer Science	6. 最初と最後の頁 177 ~ 192
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-31605-1_14	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Juliano Rabelo, Mi-Young Kim, Randy Goebel, Masaharu Yoshioka, Yoshinobu Kano and Ken Satoh	4. 巻 -
2. 論文標題 COLIEE 2019 Overview	5. 発行年 2019年
3. 雑誌名 Proceedings of the Sixth International Competition on Legal Information Extraction/Entailment (COLIEE 2019)	6. 最初と最後の頁 1-9
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Masaharu Yoshioka and Zihao Song	4. 巻 -
2. 論文標題 HUKB at COLIEE 2019 Information Retrieval Task - Utilization of metadata for relevant case retrieval -	5. 発行年 2019年
3. 雑誌名 Proceedings of the Sixth International Competition on Legal Information Extraction/Entailment (COLIEE 2019)	6. 最初と最後の頁 33-37
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Simona Frenda, Noriko Kando, Viviana Patti and Paolo Rosso	4. 巻 -
2. 論文標題 Stance or insults?	5. 発行年 2019年
3. 雑誌名 Proceedings of the Ninth International Workshop on Evaluating Information Access (EVA 2019), a Satellite Workshop of the NTCIR-14 Conference	6. 最初と最後の頁 15-22
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Masaharu Yoshioka	4. 巻 -
2. 論文標題 WC3: Wikipedia Category Comprehensiveness Checker based on the DBpedia Metadata Database	5. 発行年 2018年
3. 雑誌名 Workshop and Poster Proceedings of the 8th Joint International Semantic Technology Conference co-located with the 8th Joint International Semantic Technology Conference (JIST 2018)	6. 最初と最後の頁 98-101
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Masaharu Yoshioka and Zihao Song	4. 巻 -
2. 論文標題 HUKB at COLIEE2018 Information Retrieval Task	5. 発行年 2018年
3. 雑誌名 Proceedings of the twelfth International Workshop on Juris-Informatics (JURISIN2018)	6. 最初と最後の頁 183-193
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Masaharu Yoshioka, Yoshinobu Kano, Naoki Kiyota, and Ken Satoh	4. 巻 -
2. 論文標題 Overview of Japanese Statute Law Retrieval and Entailment Task at COLIEE-2018	5. 発行年 2018年
3. 雑誌名 Proceedings of the twelfth International Workshop on Juris-Informatics (JURISIN2018)	6. 最初と最後の頁 117-128
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Han Wen-Bin, Kando Noriko	4. 巻 -
2. 論文標題 Opinion Mining with Deep Contextualized Embeddings	5. 発行年 2019年
3. 雑誌名 Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop	6. 最初と最後の頁 35-42
掲載論文のDOI (デジタルオブジェクト識別子) 10.18653/v1/N19-3006	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Zehua Yang, Yusuke Yamamoto, Takehiro Yamamoto, Noriko Kando and Hiroaki Ohshima	4. 巻 -
2. 論文標題 Finding the connection between artifact and personal knowledge of museum visitor	5. 発行年 2019年
3. 雑誌名 Proceedings of the 2nd Workshop on Evaluation of Personalized Information Retrieval, Collocated with ACM Conference on Human Interactive Information Retrieval	6. 最初と最後の頁 1-4
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計17件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 吉岡真治
2. 発表標題 Wikipediaカテゴリオントロジー - Wikipediaカテゴリの知識工学的観点からの活用を目指して -
3. 学会等名 人工知能学会第55回セマンティックウェブとオントロジー研究会,
4. 発表年 2021年

1. 発表者名 柴 孝明, 伊藤正彦
2. 発表標題 日本酒フレーバ類似度の可視化による銘柄探索システムの開発
3. 学会等名 第5回ビジュアライゼーションワークショップ
4. 発表年 2022年

1. 発表者名 中川嵩教, 小坂橋佳晃, 吉岡真治
2. 発表標題 Wikipediaカテゴリオントロジーを利用したWikipediaページのクラス分類
3. 学会等名 2020年度人工知能学会全国大会 (第34回)
4. 発表年 2020年

1. 発表者名 高丸圭一, 木村泰知, 内田ゆず, 佐々木稔, 吉岡真治, 秋葉友良, 渋谷 英潔
2. 発表標題 東京都議会会議録における議案への賛否を表明する発言の分析
3. 学会等名 2020年度人工知能学会全国大会(第34回)
4. 発表年 2020年

1. 発表者名 木村泰知, 渋谷英潔, 高丸圭一, 内田ゆず, 乙武北斗, 石下円香, 三田村照子, 吉岡真治, 秋葉友良, 小川泰弘, 佐々木稔, 横手健一, 神門 典子, 森辰則, 荒木健治, 関根聡
2. 発表標題 NTCIR15 QA Lab-PoliInfo-2 の報告およびデータセット公開
3. 学会等名 言語処理学会第27回年次大会
4. 発表年 2021年

1. 発表者名 渋谷英潔, 木村泰知, 乙武北斗, 石下円香, 秋葉友良, 小川泰弘, 横手健一, 森辰則, 神門典子
2. 発表標題 要約の自動評価手法としてのDoc2Vec の利用
3. 学会等名 情報処理学会研究報告 (NL) 246(NL-24)
4. 発表年 2020年

1. 発表者名 石下 円香, 渋谷 英潔, 神門 典子
2. 発表標題 地方議会会議録の議論構造の分析 及び平常時と非常時における会議録の構造比較
3. 学会等名 情報処理学会研究報告 (NL) 246(NL-37)
4. 発表年 2020年

1. 発表者名 岸里遼, 伊藤正彦
2. 発表標題 VR空間での散布図要素選択における触覚フィードバックの効果の評価
3. 学会等名 令和2年度 電気・情報関係学会北海道支部連合大会
4. 発表年 2020年

1. 発表者名 立浪紀彦, 吉岡真治, 神門典子, James Allan, 伊藤正彦
2. 発表標題 ユーザの興味を反映したニュースサイトの多観点スタンス分析
3. 学会等名 言語処理学会第26回年次大会発表論文集, D5-4
4. 発表年 2020年

1. 発表者名 木村泰知, 渋木英潔, 高丸圭一, 秋葉友良, 石下円香, 内田ゆず, 小川泰弘, 乙武北斗, 佐々木稔, 三田村照子, 横手健一, 吉岡真治, 神門典子
2. 発表標題 NTCIR-15 QA Lab-PoliInfo2 のタスク設計
3. 学会等名 言語処理学会第26回年次大会発表論文集, F3-2
4. 発表年 2020年

1. 発表者名 小板橋佳晃, 吉岡真治
2. 発表標題 森羅と日本語DBpediaにおける属性の取り扱いに関する比較分析
3. 学会等名 言語処理学会第26回年次大会発表論文集, D5-1
4. 発表年 2020年

1. 発表者名 中川嵩教, 小坂橋佳晃, 吉岡真治
2. 発表標題 カテゴリの親子関係の種類に基づくWikipediaカテゴリの再整理
3. 学会等名 2019年度人工知能学会全国大会(第33回)論文集, 1K2-J-4-05
4. 発表年 2019年

1. 発表者名 中川嵩教, 小坂橋佳晃, 吉岡真治
2. 発表標題 日本語Wikipediaオントロジーの構築
3. 学会等名 人工知能学会第48回セマンティックウェブとオントロジー研究会, SIG-SW0-048-07
4. 発表年 2019年

1. 発表者名 立浪紀彦, 吉岡 真治
2. 発表標題 大規模ニュース記事データベースを用いた複数観点からの記事の対比と報道スタンスの分析
3. 学会等名 情報処理学会ICS研究会2018- ICS-192
4. 発表年 2018年

1. 発表者名 中川嵩教, 吉岡真治
2. 発表標題 知識工学者のためのWikipediaカテゴリ構造の再整理
3. 学会等名 2018年度人工知能学会全国大会(第31回)
4. 発表年 2018年

1. 発表者名 中川嵩教, 小板橋佳晃, 吉岡真治
2. 発表標題 Wikipediaカテゴリの構成要素に注目したカテゴリ階層の分析
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 楊澤華, 山本祐輔, 山本岳洋, 神門典子, 大島裕明
2. 発表標題 博物館の展示物と見学者の興味を関連付ける情報の発見
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム (DEIM2019)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	伊藤 正彦  (Masahiko Itoh)  (60466422)	北海道情報大学・情報メディア学部・准教授   (30115)	
研究分担者	神門 典子  (Noriko Kando)  (80270445)	国立情報学研究所・情報社会相関研究系・教授   (62615)	

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	アラン ジェームス  (Allan James)		University of Massachusetts, Amherst Center for Intelligent Information Retrieval Professor

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
米国	University of Massachusetts, Amherst			
カナダ	University of Alberta			