様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

令和　４ 年　 ４ 月 １８ 日現在

機関番号：１７２０１
研究種目：若手研究
研究期間：2019～2021
課題番号：１９Ｋ２０６３０

研究課題名（和文）文化財書誌の機械可読化普及を目指した低コストなLinked Data自動変換

研究課題名（英文）Low-cost and Automatic Linked Data Conversion for Machine-readable Bibliographies of Cultural Properties

研究代表者
吉賀　夏子（Natsuko, Yoshiga）

佐賀大学・地域学歴史文化研究センター・研究機関研究員

研究者番号：７０４５７４９８

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：郷土に残存する多くの古記録から内容を把握するには、当時の文語によるくずし字文書を読み解く専門性が必須となる。現在このような専門性をもつ人材は数少なく、地域の歴史や文化を解析する大きな障壁となっている。本研究では、翻刻されたテキスト文から人名、出来事名、地名など読み解きの鍵となり得る固有表現（キーワード）を可能な限り機械的に抽出する手法を開発した。研究前半では、ネット上では見つけられない地域固有の表現を市民科学の観点から、郷土資料に元々関心のある市民に抽出を依頼した。研究後半では、人手による抽出結果を基に深層学習の手法を用いた固有表現抽出を行ない、大量の固有表現を高精度に抽出する手法を確立した。

研究成果の学術的意義や社会的意義
近年、我々の身の回りで起きる出来事をデータ化し、社会課題の解決に活かすデータ駆動型社会への移行が加速している。郷土の歴史資料においても同様に単なる画像への電子化に留まらず、テキスト化・機械可読化することが、人手に余る大量の資料の解析に必要であるとの認識が高まっている。本研究では、地域の歴史を知り守りたいと思う市民の助けと機械学習の力で歴史資料を低コストに機械可読化する手法を確立した。さらに、従来のデジタルアーカイブにおけるデータ提供者と利用者の役割を超えて関係者全員が文化財データを構築していく市民科学の実践にも貢献した。

研究成果の概要（英文）：In order to understand the contents of the many historical records that have survived in the local area, it is essential to have the expertise to read and understand the handwritten, kuzushiji documents in the literary language of the era. Currently, there are only a few people with such expertise, and this is a major barrier to analyzing local history and culture. In this study, we developed a method to mechanically extract unique expressions (keywords) such as names of people, events, and places from reprinted texts as much as possible, which can be the key to deciphering the text. In the first half of the study, we asked citizens who were originally interested in local materials to extract local-specific expressions that could not be found on the Internet from the perspective of citizen science. In the second half of the research, we established a method to extract unique expressions with high accuracy by using deep learning methods based on the manual extraction results.

研究分野：人文情報学

キーワード：江戸期古記録　シチズンサイエンス　深層学習　固有表現抽出　単語分散表現　機械可読化

１．研究開始当初の背景

(1)　近年、我々の身の回りに存在する多種多様な物事をデータ化し、それを社会課題の解決に活かす動き、すなわち「データ駆動型社会」への移行が世界的に加速している。この動きは、歴史文化の分野においても近年注目されている。例えば、佐賀大学地域学歴史文化研究センターでは、「日記」や「万覚帳」と呼ばれる藩主の側回りや藩政役所の業務日誌の表題のみを集めた江戸時代の目録を翻刻してテキスト化し、「小城藩日記データベース」(https://crch.dl.saga-u.ac.jp/nikki/）で検索可能にした。目録には、藩の行政のみでなく領地、農民、冠婚葬祭などに関する多様な出来事が元の日記を要約する形で時系列に記載されており、当時の行政、文化の調査および研究に役立つものとなっている。しかし、全目録数は当初10万近くに上ると予想されるうえに、漢字が95％以上を占め、当時の文語体（候文）で記載されている。そのため、テキスト化された短い目録文であっても、専門家でないほとんどの一般市民には何が書かれているのか把握することが難しいのが現状であり、その利活用の広がりは限定的にならざるを得ない。

(2)　日記目録中の記事文から機械可読な形で人名、地名、出来事、職名など大まかな意味を付与した文字列としてラベル付け（表1）することができれば、市民は記載内容を理解することが容易になることに加え、記載内容を機械的に定量分析できるはずである。言い換えると、従来の限られた研究者個人が長期にわたり資料を読み込んで考察する手法に加えて、考察の裏付けとなり得るデータを世界中から収集し、第三者にも明確に説明可能な分析結果を導きだす環境を提供できることになる。研究者以外の郷土史に関心のある市民にとっても、原文の解読（翻刻）経験がなくともアーカイブを通じて大まかに記載内容を考察するきっかけを与えることになる。しかし、目録記事文を機械可読のデータに加工するのは容易ではない。なぜなら、(1) で述べた通り、記載内容を機械可読化するにも、翻刻同様に単語に対するラベル付けに専門知識が必須である上に、数十万以上の単語に対して手作業や簡単な機械的処理のみでラベル付けを進めることは非現実的であるためである。

表1　日記目録における固有表現クラスの一覧

| 固有表現クラス名 | 説明 |
|---|---|
| EVENT / 出来事 | 出来事の名称 |
| TERMS / 候文用語 | 接続詞，定型句 |
| ROLE / 役職・役割 | 役職，家族関係 |
| PERSON (JINMEI) / 人名 | 人名，呼称 |
| PLACE / 場所 | 地名，建物の呼称 |
| QUANTITY / 数量 | 数および単位を表す語 |
| DATE / 日時 | 日時を表す語 |

(3)　これまで申請者は、非構造化データである文化財書誌に対し、その構造を含めて機械可読なデータに変換するため、自然文で書かれた書誌からキーワードを形態素解析ツールを用いて抽出し、人名、地名といった大まかな意味のラベル付けを行なった。その上で、図1に示すようなLinkedDataのような機械可読の構造化データへ低コストで自動変換する手法を開発してきた。この手法における課題は、ラベル付けされたキーワード（以下、固有表現と呼ぶ。）を抽出するための準備作業（図1右上の形態素解析ツールによる固有表現抽出（ユーザ辞書データの作成）は手作業であり、一定の個人負担が求められる点である。例えば、専門用語や定型句は書誌データ全体を通じて使用されるため形態素解析のような機械的処理には有効である一方で、人名は書誌データが増補される度に似たような人名が多く出現するにもかかわらず、登録済み単語と1文字でも違えば新たに追加登録する必要がある。例えば、「中嶋神左衛門」と「中嶋金左衛門」のような違いである。人ならば容易にできることが機械的処理を実行する際は大きな障壁となっている。

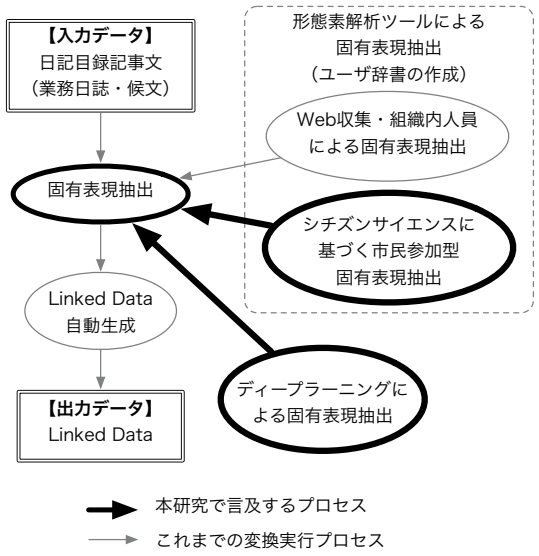


図1　申請者による低コストなLinkedData（機械可読）化システムの概要図

２．研究の目的

本研究の主目的は、日記目録のような大量の自然文を含む文化財書誌データを可能な限り個人負担を減らして低コストに機械可読化することである。現状、申請者が開発した固有表現抽出システムを用いて一定の機械的処理は可能であるが、それには書誌中の自然文を構成する単語を抜き出して、目録を読むための文法用語や人名、地名などの地域特有の固有表現をあらかじめ辞書

として大量に収集する必要がある。この課題を解決するために、2 つの目標を設定した。

① 学術研究に市民が積極的に参加するシチズンサイエンスの枠組みを通じ、郷土資料の読み解きを実際に行える市民と協働し、Web などからの自動収集が困難な対象文化財の固有表現を抽出するシステムを構築する。

② 市民と共に収集した固有表現を基にしてディープラーニング（深層学習）を行い、未知の目録記事文から固有表現を類推して自動抽出する技術の開発を行う。

３．研究の方法

(1) シチズンサイエンスに基づく固有表現抽出システムでは、手作業で固有表現抽出タスクを実行する参加者、その支援ソフトウェア、参加者による抽出結果の質について留意し、固有表現抽出の品質を損ねないための工夫を行なう必要がある。

① タスク参加者は、江戸期の文語体「候文」で書かれた翻刻済みテキストを読み、地域の出来事、候文用語、役職・役割、人名、場所、数量、日時の 7 つのクラスを意味する固有表現を抜き出し、各クラスのラベル付けを行う。そのため、候文を読める大学の人員に加え、郷土の知識になじみがあり、一般的な江戸期の古文書の読解に習熟した 60 代から 80 代の地元市民にタスクを依頼した。最終的に、記事文 4 万件の固有表現抽出を 3 名無償、5 名有償の計 8 名でタスクを実行した。

② 支援ソフトウェアの構築においては、図 2 で示す通り、フロントエンドのウェブアプリケーションであるタスク支援ソフトウェアを、バックエンドのシステム管理機能の下で稼働させる構成をとっている。タスク参加者および後述の管理者は、ダッシュボードと呼ぶ固有表現抽出済みの記事文総数および月別総数や個人のタスク完了記事文数などを確認できる画面で進捗状況を把握する。タスク遂行画面では、未解析の目録記事文のひとつがバックエンドで決定され、ランダムかつ他のタスク参加者と重複を避けて表示される。

表示した記事文では、あらかじめ形態素解析ツール MeCab とそのユーザ辞書を用いて自動的に固有表現抽出が行われており、その抽出結果をタスク参加者が確認し、修正点があれば正しい単語と対応する固有表現ラベルを登録する。バックエンドでは、タスク参加者のログイン日時、操作内容、固有表現抽出の進捗を記録し、データベースに蓄積する。なお、本ソフトウェアを使用する前には、タスク参加者に対して操作説明会を開いて画面操作や固有表現抽出に必要なルールの具体例などを共有し、後に参加者が単独でタスクを行う際の疑問点や不安を解消した。

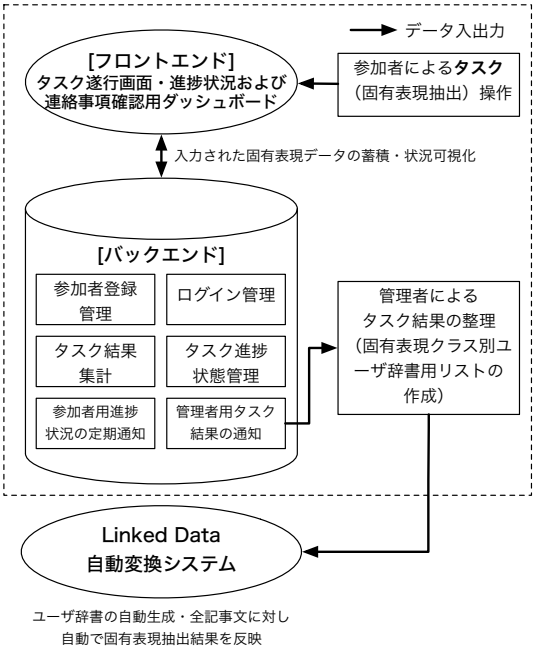


図 2 タスク支援ソフトウェア（点線内）の概要図

③ 参加者による抽出結果の質の担保

タスク参加者には一定の候文読解能力があるが、当然ながら解釈の仕方に個人差が生じたり誤りが発生する余地がある。本研究では、タスク参加者の抽出結果を直接正解データとして用いず、バックエンドから出力されたクラス別固有表現データを随時整理する管理者を設置した。管理者は、抽出された固有表現が妥当なクラスに仕分けられているかを目視で確認した後、確認済み固有表現データを LinkedData 変換（機械可読化）システムに投入する。このシステムは、新しい辞書を自動的に再構築し、その辞書を用いて ② でタスク参加者が参照する基本の固有表現抽出結果をアップデートする。

加えて、タスク参加者および管理者は、最新の抽出状況を参照できるウェブサイト「固有表現リスト」（https://winter.ai.is.saga-u.ac.jp/cs/ne-words.php）にて、固有表現がどのクラスに何回判定されたかを参考にできる。例えば、タスク参加者が「肥州」という語が職名を呼び名に転じた人名であるのか、単なる場所を表す語なのかクラス判定に迷う場合、当サイトで「肥州」を検索すると、これまでに判定された結果を実際の目録記事文を参照しながら確認することができる。以上の仕組みを用いてタスク参加者が抽出した固有表現データの質を担保した。

(2) (1) のシチズンサイエンスで収集した固有表現抽出手法は、対象記事文が 2、3 万件を超え、ユーザ辞書がアップデートされるにつれて、背景 (3) で述べた通り、ユーザ辞書に未登録の新語

は固有表現抽出できないことと、タスク参加者間でのクラス判定に個人差が生じることで、ユーザ辞書の管理者による内容の整理に手間取るというデメリットを顕著にした。
そこで、申請者は、単語分散表現（Word embeddings）と呼ばれる、単語間の類似度を高次元の実数ベクトルで数値化して表現するディープラーニング手法を用いて実行することに着目した。しかし、手持ちの江戸期の固有表現クラスラベル付きの目録記事文のみでは、基本的に大量の学習データを必要とするディープラーニングのモデル構築には不足する。そのため、目録記事文の候文のような江戸期の文語体であっても、文中に出現する単語や文字およびそれらに付随する意味（固有表現クラス）が現代日本語と大きくかけ離れていなければ、汎用の現代日本語コーパスから生成した単語分散表現に、ドメインに特化した少数の目録記事文を用いて追学習を行うことで、高精度な固有表現抽出結果が得られると仮定した。

また、(1)③ で述べたように、目録記事文中の単語には、文脈に応じて固有表現クラスが変化するものがあるため、文脈を考慮した単語分散表現を構築する必要がある。本研究では、以上のような要件を満たすために、単語分散表現の構築に文脈を考慮可能な Attention 機構 [1] を採用している Flair フレームワークおよびそのライブラリ [2] を用いて、文字レベルの単語分散表現を組み合わせた学習モデルを構築し、目録記事文に対する固有表現判定モデルを生成した。
追学習および精度判定に用いる文は、目録記事文の登録番号 1 から 40,000 までの文からランダムに 5,000 件刻みで必要件数を選択しデータセットとしたものである。毎回ランダムに作成された各データセットで 3 回ずつ判定モデルを構築して、適合率、再現率および F 値の測定を行なった。F 値などの抽出精度の測定は、事前学習と追学習（Fine-tuning）で固有表現判定モデルを構築する際に Flair 組み込み機能で行った。

## ４．研究成果

本研究では、市民参加による固有表現抽出の確認作業を 2019 年 5 月から 2020 年 5 月まで行った。最終的には 2020 年 10 月に全目録記事 73,984 件の翻刻された目録記事文のデータベース登録が完了した。登録完了時点で抽出内容確認済み 40,000 件の目録記事文を基準の正解データとし、過去の固有表現抽出結果と比較することで、抽出精度の指標である適合率（Precision）、再現率（Recall）および F 値（F-measure）を測定した。

### (1)　市民参加によるクラス別固有表現抽出評価結果

① 市民参加によるタスク実行期間内での MeCab ユーザ辞書のデータ追加および整理作業の結果として、登録完了時点での抽出済み固有表現の出現数、出現数の割合および固有表現数を表 2 に示す。また、各クラスにおける固有表現を出現度数順に並べた際の出現度数から算出される占有率（分布）を求めた結果、クラスによって分布の状況に特徴が見られた。例えば、人名クラス（図 3 左）の場合、固有表現出現数の上位 100 位までででこのクラス全体の 26.4%を占め、1,000 位までで 65.9%を占有することを示した。

表 2　市民参加型タスク終了時（2020 年 10 月時点）の記事文 73,984 件に出現した固有表現の出現数、全固有表現の出現数に対する各クラスの占める割合、および固有表現数

| クラス名 | 固有表現の出現数 | 出現数の割合（%） | 固有表現数 |
|---|---|---|---|
| 全クラス | 409,004 | 100.0 | 18,141 |
| EVENT / 出来事 | 131,032 | 32.0 | 5,781 |
| TERMS / 候文用語 | 117,384 | 28.7 | 257 |
| PERSON / 人名 | 52,098 | 12.7 | 7,238 |
| ROLE / 役職・役割 | 50,607 | 12.4 | 2,019 |
| PLACE / 場所 | 37,685 | 9.2 | 1,897 |
| QUANTITY / 数量 | 10,931 | 2.7 | 809 |
| DATE / 日時 | 9,267 | 2.3 | 581 |

これに対し、候文用語クラス（図 3 右）では、上位 1 位の辞書登録でこのクラスの全出現数の 30.1%、上位 100 位までの登録で 98.8%をそれぞれ占有した。つまり、各クラスにおける占有率は、頻出の固有表現の個数が上昇するにつれて単調増加することから、未知語のユーザ辞書登録は正確な固有表現クラスの判定に寄与することは間違いないが、人名および出来事クラスについては、固有表現が他のクラスに比べて多種であり、精度向上を図るには限界があることが示唆された。

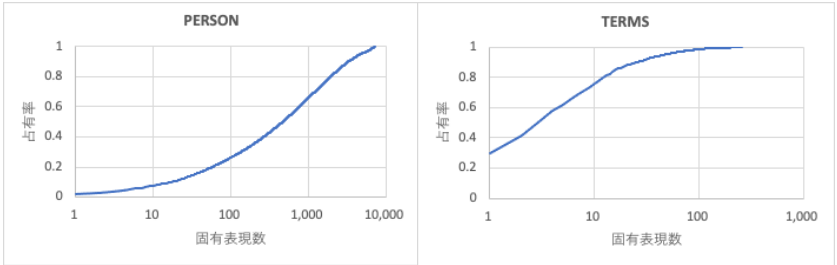

図 3　人名（PERSON）および候文用語（TERMS）クラスの固有表現数に対する占有率

②1 年間のタスクを通じて固有表現抽出評価結果の推移を調査したところ、タスク実行初期および固有表現参加者の抽出に対する解釈の違いが個人間で拡大したため、抽出精度の一時的な低下が見られた。

その理由は、クラス間および個人間で抽出難易度に差があるためである。人名と候文用語の各クラスでの抽出精度は他のクラスより常に高かったが、全記事文中に占める固有表現の出現数が 3 割を超える出来事クラスについては、2019 年 9 月以降に再現率が適合率を上回り、参加者の正解率（図 4）では人名や候文用語の次に高い結果を示した。役職・役割、数量、日時および場所については、これらの順で正解率が低下した。一般に、これらのクラスに属する語は多義的であり、文脈から判定せざるを得ない場合が多く、所定のクラス判定のルールのみでは判断が難しい。また、「御」のような冠詞や「様」などの接尾辞を含むか含まないか、複数の単語が連なった固有表現をどこまでまとめるかなど、文中のどの部分を固有表現として抽出するか判断が難しい語が多く含まれている。その一方、これらのクラスは適合率が再現率を上回っていた。これは、クラス判定は不正解だったものの固有表現の抽出そのものはできている場合が多いことを示している。さらに、抽出すべき固有表現の種類数自体は、役職・役割、数量、日時および場所においてそれぞれ 2,019、809、581 および 1,897 であり（表 2）、人手での抽出は量的にも困難ではない。したがって、クラス判定ルールをより明確にして辞書による再利用効果を高める、あるいは個別での対処に注力するなどの工夫で抽出精度をさらに向上させる余地があると考えられる。

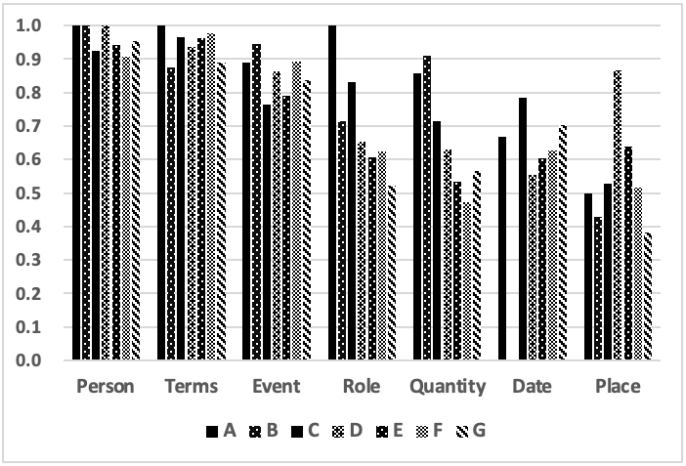

図 4　固有表現クラスおよび参加者別タスク正解率。A および B：無償、C から G：有償

(2)　ディープラーニングによるクラス別固有表現抽出精度の検証

データセットが 5,000 件から 40,000 件に増加するに伴い、全クラスの F 値が向上し、その最小値と最大値の差が小さくなった。また、本研究で高精度と定義している 0.9 以上の F 値を超えるために必要な記事文数は、表 3 に示す通り、クラスによって差異が見られた。

具体的には、人名および候文用語クラスは、5,000 件の学習で F 値 0.9 以上を示した（表 3）。これらのクラスの次に、場所、出来事、役職・役名の順で、20,000 件程度の学習により高精度な結果を得られた。対して、日時および数量クラスでは、追学習データを記事文の半数以上である 40,000 件を投入することで F 値が 0.9 を超えた。最終的に、2020 年 10 月時点までに記事文から抽出した全固有表現数 409,004 に各クラスが占める割合を考慮すると、25,000 件（全記事文の約 33.8%）の正解データセットがあれば、表 2 で示した全固有表現の数で 95.0%を占める人名、候文用語、場所、出来事、役職・役割の各クラスで、F 値 0.9 を超えると示唆される。

表 3　ディープラーニングで適合率、再現率および F 値が 0.9 以上になるために必要な追学習用データセット数

| | 適合率 | 再現率 | F 値 |
|---|---|---|---|
| TERMS / 候文用語 | 5,000 | 5,000 | 5,000 |
| PERSON / 人名 | 5,000 | 5,000 | 5,000 |
| PLACE / 場所 | 15,000 | 15,000 | 15,000 |
| EVENT / 出来事 | 20,000 | 15,000 | 20,000 |
| ROLE / 役職・役割 | 25,000 | 20,000 | 25,000 |
| DATE / 日時 | 40,000 | 40,000 | 40,000 |
| QUANTITY / 数量 | なし | 30,000 | 40,000 |

以上の結果から、地域の固有表現を多数含む目録記事文のような古記録からの固有表現抽出には、研究機関の枠組みを超えた市民の協力が必要である。さらに、市民の協力で得られた学習データセットを用いて生成した単語分散表現の利用は、従来の形態素解析用辞書を基にした固有表現抽出手法の課題を解決する手がかりとなることが明らかになった。[3]

*参考文献

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, Vol. abs/1706.03762, , 2017.

[2] Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. FLAIR: An easy-to-use framework for state-of-the-art NLP. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 54–59, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[3] 吉賀夏子, 堀良彰, 只木進一, 永崎研宣, 伊藤昭弘. 郷土に残存する江戸期古記録の機械可読化を目的とした市民参加および機械学習による固有表現抽出. 情報処理学会論文誌, Vol. 63, No. 2, pp. 310–323, feb 2022.

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3件／うち国際共著　0件／うちオープンアクセス　0件）

| | |
|---|---|
| １．著者名<br>吉賀 夏子 | ４．巻<br>736 |
| ２．論文標題<br>地域市民と創出する文化財データ | ５．発行年<br>2020年 |
| ３．雑誌名<br>考古学ジャーナル | ６．最初と最後の頁<br>27-28 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| １．著者名<br>吉賀 夏子，只木 進一 | ４．巻<br>2019 |
| ２．論文標題<br>低コストなLinked Data化を目指したクラウドソーシングによる固有表現収集の試み | ５．発行年<br>2019年 |
| ３．雑誌名<br>じんもんこん2019論文集 | ６．最初と最後の頁<br>239-244 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| １．著者名<br>吉賀 夏子、堀 良彰、只木 進一、永崎 研宣、伊藤 昭弘 | ４．巻<br>63 |
| ２．論文標題<br>郷土に残存する江戸期古記録の機械可読化を目的とした市民参加および機械学習による固有表現抽出 | ５．発行年<br>2022年 |
| ３．雑誌名<br>情報処理学会論文誌 | ６．最初と最後の頁<br>310～323 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.20729/00216238 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

〔学会発表〕　計4件（うち招待講演　3件／うち国際学会　1件）

| |
|---|
| １．発表者名<br>吉賀 夏子，堀 良彰，永崎 研宣 |
| ２．発表標題<br>候文における文字単位の単語分散表現モデルに基づく固有表現抽出手法 |
| ３．学会等名<br>研究報告人文科学とコンピュータ（CH） |
| ４．発表年<br>2021年 |

１．発表者名
吉賀夏子

２．発表標題
小城藩日記プロジェクトの紹介

３．学会等名
第122回人文学とコンピュータ研究会（情報処理学会）／第13回地域学シンポジウム（招待講演）

４．発表年
2020年

１．発表者名
吉賀夏子

２．発表標題
低コストな文化財書誌の機械可読化を目指して

３．学会等名
人文学とコンピュータシンポジウム2019 企画セッション「若手研究者によるCH／人文情報学」（招待講演）

４．発表年
2019年

１．発表者名
Natsuko Yoshiga

２．発表標題
Publication and Utilization of IIIF Images in the Ogihan Nikki Database

３．学会等名
IIIF Week: Japan Showcase Session（招待講演）（国際学会）

４．発表年
2020年

〔図書〕　計1件

| １．著者名 | ４．発行年 |
|---|---|
| The National Museum of Japanese History　Makoto Goto, Satoru Nakamura, Chifumi Nishioka, Arianti Ayu Puspita, Taizo Yamada, Yuta Hashimoto, Natsuko Yoshiga, Tatsuki Sekino, Naoki Kokaze, and Shohei Yamasaki | 2021年 |
| ２．出版社 | ５．総ページ数 |
| University of Michigan Press | 216 |
| ３．書名 | |
| Japanese and Asian Historical Research in the Digital Age (Conversion of Historical Ogihan Business Records into Linked Open Data via Human-Machine Cooperation) | |

〔産業財産権〕

〔その他〕

小城藩日記データベース
https://crch.dl.saga-u.ac.jp/nikki/

固有表現リスト
https://winter.ai.is.saga-u.ac.jp/cs/ne-words.php

小城藩日記プロジェクト - UDC2019 NO.188
https://winter.ai.is.saga-u.ac.jp/udc2019/

Named Entity Recognition (NER) for Ogihan Nikki Mokuroku titles（学習モデル生成プログラム）
https://colab.research.google.com/drive/1vrxCIx2o-4GiP8zewZO18UngEAQ6Lfnv?usp=sharing

Named Entity Recognition (NER) for Ogihan Nikki Mokuroku titles（固有表現クラス判定プログラム）
https://colab.research.google.com/drive/1_30rwEPSp6P5EOLzFn_g1xN-wEX-Dw4I?usp=sharing

## ６．研究組織

| | 氏名<br>（ローマ字氏名）<br>（研究者番号） | 所属研究機関・部局・職<br>（機関番号） | 備考 |
|---|---|---|---|
| 研究協力者 | 伊藤　昭弘<br>(Ito Akihiro) | | |
| 研究協力者 | 堀　良彰<br>(Hori Yoshiaki) | | |

## 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 |
|---|---|