Vol.2016-CH-109 No.3
2016/1/30

古典籍書誌注記文からの作品構造および関連人物のつながりを明らかにする周辺情報抽出

吉賀 夏子[1,a)]　渡辺 健次[2,b)]　只木 進一[1,c)]

概要：江戸時代を主とした古典籍の研究では，書名，編著者，刊行年等の通常の書誌情報以外に，表紙，見返し，序跋文等，書籍の部品に関する多くの周辺情報が必要である．これらの情報は，通常の書誌では注記に記載されている．本稿では，こうした情報を半自動的に抽出する方法について提案する．事前に，書名や編著者名等の対象書籍に関連する固有情報とともに，国立国会図書館等で公開されている典籍情報を取得し，ユーザ辞書として保有する．また，辞書中に出現しない人名を抽出しやすくするために，注記中にそれらの人名が出現するパターンの推論規則を定義する．このような準備の下，注記に対して形態素解析を行う．このような手続きを行うことにより，注記より，日時，地名，人名，部品名，ロール名（人物が出版に果たした役割）を含む情報を高精度で抽出することが可能となる．提案手法の具体的なコレクションに対する実行結果を示し，その精度と課題について議論する．

キーワード：書誌記述, 正規化, 古典籍, 周辺情報, 書誌データの自動作成

Extracting Peripheral Information from Bibliography Information of Historical Rare Books to Clarify Those Structures and Related People.

Yoshiga Natsuko[1,a)]　Watanabe Kenzi[2,b)]　Tadaki Shin-ichi[1,c)]

Abstract: On researches about historical rare books mainly published in Edo era, the researchers need peripheral information such as book covers, prefaces, illustrations and so on, to understand the formation of the books. For realization of this concept by using bibliographical data, a framework to clarify events that people or things connect to the books are expected to effectively share bibliographic data on the Web. On the other hand, actual constructions of the data based on the framework by hand need a great deal of labor, especially, at the point of extractions and identifications about keywords which express the events to understand the formation of the books. In this article, we designed a tool to extract the keywords from miscellaneous fields of existing bibliographical data to assist data creators who want to change relational data into structured data as Linked Data. Also, we reported the actual method and effectiveness of the tool. In result, peripheral information about bookbindings and roles are essential to extract unknown human names in the rare book. We need more bibliographical data as sample to prove about the result.

Keywords: bibliographic description, normalization, historical rare book, peripheral information, automation of metadata

1. はじめに

近年，国立国会図書館デジタルライブラリー [1]，古典籍総合目録データベース [2] をはじめとして，明治頃以前に国内で刊行あるいは書写された資料である「古典籍」の Web 公開が進んでいる．例えば，[2] に集められている書誌目録

[1] 佐賀大学大学院工学系研究科
Graduate School of Science and Engineering, Saga University
[2] 広島大学大学院教育学研究科
Graduate School of Education, Hiroshima University
[a)] 15634011@edu.cc.saga-u.ac.jp
[b)] wtnbk@hiroshima-u.ac.jp
[c)] tadaki@cc.saga-u.ac.jp

数は，2015 年 11 月の時点で約 542,800，著作としては約 474,900 である．書籍の比較や関係を調査するために，書籍の来歴や関係者などの周辺情報を，オンラインで公開されている膨大な情報の中から探し出すことは困難である．

このような困難の原因の一つは，誰が，いつ，どこで，どのような役割を果たしたかという周辺情報が，従来の書誌情報では注記という構造化されていない部分に記載されてきたことである．この部分を構造化し，効率的に検索可能とすることにより，書誌情報の一層の活用が可能となる．

本論文で述べる「周辺情報」とは，目録中に出現する全ての時，場所，人名，専門用語等，古典籍研究を行う手がかりとなり得る固有表現である．例えば，表 1 において，書名の「(開化新作）大津絵ぶし初号」，書名末尾に付された，外題を表す「(外)」，刊行・書写年次の「(明治期)」等は周辺情報である．注記項目中の「奥付」「(地本錦絵）問屋□□□□辻□屋文助」等，資料中のある箇所について詳細を記述した際に出現した表現も周辺情報に該当する．

表 1 佐賀大学附属図書館所蔵「市場直次郎コレクション」（http://www.dl.saga-u.ac.jp/OgiNabesima/）の書誌データの一例（□は汚損等で解読不能であることを示す）

整理番号	137
分類1	文学
分類2	近世歌謡
分類3	総記
書名	（開化新作）大津絵ぶし 初号（外）
読み	おおつえぶししょごう
書型	中
巻冊	1巻1冊
編著者	-
刊行・書写年次	（明治期）
西暦	1868
刊写	刊
注記	奥付は刷りが悪く判読できない部分もあるが、「（地本錦繪）問屋□□□□辻□屋文助」。表紙は刷り表紙で、「本を読む女」図。
画像	http://www.dl.saga-u.ac.jp/OgiNabesima/ohtsuebushi/137/
画像公開	公開
印記	1
文庫	市場
他	5

研究者はこの周辺情報と実際の資料から得られる情報を照らし合わせながら，元々の研究テーマにつながる資料を探索することで，参照資料間の位置付けを把握する．仮に，このような情報がオンラインで機械的に取得可能であれば，研究者の要求に応じた様々なデータ解析が可能となるはずである．そこで，近年は，書誌項目（属性）を構造とともに Web 公開することで，検索エンジンあるいは独自のアプリケーション等が機械的に探索可能な公開様式である Linked Data の考えに沿って書誌を構造化データにする取り組みが進んでいる．

しかし，周辺情報取得のための書誌記述およびデータ作成には，いくつかの課題がある．このうち，本研究では，1) 書誌項目中のデータに対し，古典籍研究者が求める書誌記述と機械的なデータ取得に対応した正規化が不十分である，2) 既存の表形式データを正規化した上で構造化することは手間がかかる，という点に着目する．

我々は，課題 1) については，これまでに古典籍の成立を分析するために必要な概念モデルを提案してきた [3][4]．さらに，実存する古典籍書誌データをその概念モデルに投入して機械的データ取得の有効性を確認した．この概念モデルは既存の書籍および文化財用の形式オントロジーを基に拡張したドメインオントロジーである．提案したオントロジー構築の目的は，古典籍の成立を把握するために，表紙，序跋等の製本用語で書誌記述を部品化することである．書誌記述が部品化され，実際に書誌データが部品ごとに構造化されると，部品中に出現する概念である，時，場所，人名等の固有表現に関して機械的取得が容易になる．

一方，課題 2) で述べた通り，上記の書誌記述に，表形式で作成された既存の書誌データを対応させることには手間が掛かる．何故なら，人手で OpenRDF[5] などの変換ツールを用いながら，固有表現がどの書誌項目，あるいは部品のデータか整理し，可能であれば URI に対応づける作業が必要なためである．特に，書誌の注記文からの固有表現取得は，データ作成者は専門用語や人名，著作名等に関する予備知識が求められるため，固有表現抽出に手間が掛かる．

本研究では，この作業を支援するために，まず，古典籍で必要な固有表現とは何かを整理する．次に，現代語で単純な文章で記述されている古典籍の注記文から，日時および時節，地名，人名，著作名，部品名を含む専門用語等の固有表現の候補を高精度に抽出し，構造化データに変換することを支援する半自動化ツールを考案する．そして，佐賀大学附属図書館所蔵の「市場直次郎コレクション」[6] 書誌データ [7]（以下，*ichiba* と呼ぶ.）を用いて，固有表現の抽出を試みる．その際に得られた結果と考察を報告する．

2. 周辺情報共有の観点からみた既存の書誌記述の問題点

2.1 注記文を含めた書誌の正規化

古典籍の書誌データの多くは，書籍として書名，責任表示（著者名），刊行年等の現代書誌と同様の固定された項目構成で記述されている．しかし，国文学研究資料館らが策定している目録記述規則によると，これらの項目に加え，古典籍の検索の際，有用となる項目は，形態，書籍同士の親子関係，さらに，注記の情報が研究に有用とされている．注記には，奥書の文書，表紙の作者，印刷様式，書肆名（出版者）などの特徴が，記述形式に強い制約のない自然文で記載されている．有用であるにもかかわらず，正規化されていないため，そのデータは機械的に取得できない．

そこで，注記文においても周辺情報を取得可能な記述を考慮した上で正規化を行う必要がある．本論文で提案する正規化とは以下の通りである．

(1) 書籍を記述する情報を，属性と値の組にする．各属性は書誌に凡例があればそれを用いる．各属性に対し，言葉の意味を明確に定義した上，URI を識別子として発行し，Web で一意に参照可能とする．さらに，各属性には適切な型を持つ値を設定する．特に，注記に含まれていた部品に関する書誌情報を属性と値の組として取り出す．
(2) 古典籍書誌データの採録では，資料の解読からは値が確定されず，データ採録（提供）者の注釈・解釈が含まれていることがある [8]．そのため，書誌データに「？」，括弧，「〜か？」等の注釈的記号・記述が含まれていることがある．これらの注釈的記号・記述を値から分離する．
(3) 人名および分類名に対して，表記ゆれを吸収するため，日本国内であれば，国立国会図書館などの公的機関で発行されている典拠データ [9] の標目 URI を紐付ける．

2.2 既存の表形式書誌データを構造化するためにかかるコスト

2.1 節にて述べた正規化を，実存する表形式書誌データ *ichiba* に手作業で適用したところ，属性の設定，固有表現の同定，URI の付与，データ変換等，多くのコストを要した [3]．このような事は，[10] で Linked Data 化の普及が遅れる要因と指摘されている．しかし，一度書誌データの Linked Data 変換が完了し，公開が行われると，その書誌データは Web を通じて機械的に活用可能となる．

そこで，書誌中の周辺情報取得に必要な固有表現を抽出し，データ提供者の構造化（Linked Data）化支援を行うツールを提案するのが本論文の目的である（3.2 節）．固有表現の抽出技術に関しては，これまで [11] 等，多数の研究がなされている．これらの研究では，日本語の自然文から単語を可能な限り正確に分かち書きで抽出する．その後に，抽出単語がどのような固有表現クラス（種別）であるかを対象自然文の用途から探り，タグ付けを行う．このような固有表現の解析により，文脈を理解し適切な応答を返すことが可能となる．現在は，質問応答技術 [12] 等の基礎となっている．

提案する手法で取り扱うデータは古典籍であり，最終的に人の目による確認は必須である．提案する手法での固有表現の抽出精度が実用的であれば，その後のデータ構造化は自動で実行可能となる．さらに，注記内の自然文から直接属性と値の組を取り出せること，データの更新および修正が容易であること，固有表現の抽出に形態素解析器を用いるため，関連知識を収集したユーザ辞書を利用可能であることも利点である．本論文では，特に，提案する手法を *ichiba* 注記文に対して実行した際の，固有表現の抽出精度について述べる．

3. データ構造化手法

3.1 固有表現の定義

本研究では，文中の単語が属する概念クラスを固有表現と呼ぶ．

- 時（年号および時節の用語）
- 場所（出版地）
- 人名および団体名
- 部品名（本研究で独自に決めた，古典籍を構成する物理および論理的な構成について呼称する用語群．表紙，序，跋，見返し等.）
- ロール名（古典籍成立に関与した人物が担う役割を指す名称．編，著，画，写等.）
- 著作名（主に鉤括弧で記述されている固有名詞）
- 部品名およびロール名以外の分野に特化した専門用語
- 分類名（後印本，刊本，写本，無名氏，洒落本等）

これらの中で，研究者が周辺情報として求める固有表現とは，時，場所，人名，部品名，ロール名，著作名である．その理由は，古典籍の書誌記述では，書籍の内容あるいは形状や保存状態等の客観的な情報に加えて，「誰が，何時，何処で，何をした」という出来事，すなわち，成立状況を把握する手がかりを得るための情報が求められている．

文化財分野の形式的な概念モデルとして，CIDOC CRM[13]，EDM[14]，ミュージアム資料情報構造化モデル [15] が知られている．これらのモデルは，いずれも出来事中心に文化財の構成要素を分析可能な書誌記述の実現を目的としている．これらの概念モデルを基に実用的な書誌記述を考案し，その書誌記述に沿った，Web で参照可能なデータを構築すれば，ある古典籍の成立に，ある人物や物事が，何時，何処で，どのように関与したのかが機械的に取得可能になる [4].

3.2 データ構造化手法

3.2.1 注記文の概要

提案手法の主な目的は，主に注記文から 2.1 節で示す正規化を行い，書誌データの構造化を支援するため，研究者に必要な固有表現を抽出することである．固有表現を抽出する注記文のサンプルとして，*ichiba* の注記文（表 2）を用いた．*ichiba* の巻数を考慮しない著作としての書誌レコード数は 222 である．また，注記文は，自然文であり構文は明確に定まっていない．しかし，注記は古典籍の特徴を記述する目的で設置されている．したがって，構文は単純であり，一文に含まれる書誌項目（以下，属性と呼ぶ.）およびその値の組数は 1-10 程度である．

表 2　*ichiba* の注記文例

書名	注記
青楼心得艸	安政四年一月蓬莱山人序．天明五年刊「息子部屋」ノ改題本ナリ．
傾城買二筋道	再版．寛政十年春式亭三馬序． 雪華画．寛政十年一月梅暮里谷峨序． 梅暮里谷峨跋．
【娼婦教導】花街風流解	栗本伊賀丸序．

3.2.2　形態素解析器 Mecab 用ユーザ辞書の作成

固有表現の抽出には，Web に導入や使用例が多い形態素解析器 Mecab[16] を用いる．

その辞書データには，基本辞書として UniDic[17] を採用する．そして，基本辞書とは別に，ユーザ辞書を作成する．ユーザ辞書には，時節用語（二十四節気および雑節 [18]，月の名称および別名 [19]，十干十二支 [20]，年号 [21]），人名典拠 [9] の標目および別名データを追加する．また，部品名，ロール名，書誌学用語等の専門用語を，古典籍総合目録データベース [2] の「利用のしかた」[22] および *ichiba* の凡例から抽出し追加する．さらに，*ichiba* の書名，編著者名をキーワードとして自動的に収集した関連書誌データから，固有表現に相当する候補単語を抽出し追加する（表 3）．

表 3　固有表現クラス別の候補単語数

固有表現クラス	単語数	単語の例
（参考）Mecab 基本辞書	756463	（Unidic 登録単語）
時（時節）	1322	安政，四年，春，卯月，己巳
地名	0*	江戸，京都，東京，大阪
人名（団体名）	854322	蓬莱山人，梅暮里谷峨，栄久堂
部品名	63	表示，序，跋，見返し，奥書
ロール名	34	作，画，翻訳，編，写，述
著作名	556	息子部屋，青楼心得艸，八艘飛

*Mecab 基本辞書を利用

Mecab ユーザ辞書上の各登録単語データには，固有表現クラス名を追加し，解析結果でユーザ辞書の使用状況を確認可能とする．また，Mecab 辞書の変換コスト値は公式サイトに示す方法 [16] で自動で作成し，独自の変更は加えていない．

3.2.3　解析スクリプト

以下の手順で各固有表現クラスに関係する単語を抽出し，固有表現クラス名，レコード中の位置情報等を紐付け，リレーショナルデータベースに中間データとして格納する（図 1）．

(1) 各レコードの注記文を句読点（．および，）で分解し，可能な限り，一文に含まれる意味を単純化する．
(2) 各文に対し Mecab による形態素解析を行う．
(3) 解析結果を一旦データベースに全解析データとして格納する．解析データは，Mecab で検出した単語ごとに，レコード番号，文番号（注記文を句読点で分割した後の各文に対し割り当てた識別番号），開始位置，終了位置，単語，Mecab 解析結果で構成されている CSV データである．このデータは後出のデータクリーニングで使用する．
(4) 解析結果に固有表現抽出スクリプトを適用する．このスクリプトはモジュールとなっていて，各モジュールは時および数詞（年月日を抽出しアラビア数字表記に統一するため），鉤括弧中および *ichiba* での「○」の後に続く著作名，部品名・ロール名・専門用語，人名，地名，未知語に関する単語を，解析済み注記文をスキャンして抽出する．なお，モジュール実行は抽出が容易な，時，部品名等のスキャンから始め，抽出済み単語に対しては可能な限り除外してスキャンを継続する．注記文中で抽出が難しい固有表現は，地名および人名である．人名判定の際，屋号（一門あるいは家柄等の特徴を表す称号 [23].）が付加された未知語である場合は人名と判定する．
(5) 未知語の抽出モジュール（以下，人名推定モジュール）では，部品名およびロール名の前後にある未知語について，それらの位置確認と文字種判定を行い，人名と推定可能な未知語を人名と判定する．例えば，図 2 で「享和四年春原のきつね序」という注記文では，部品名「序」が文末に位置している．この場合，「享和四年」は事前に時の固有表現と判定されているため，未知語である「春原のきつね」の部分は人名と推定可能である．よって，図 2 の A では，結果的に不正解であるが，「春原のきつね」は人名と判定される．
(6) 固有表現クラス名が判明した単語について，「固有表現クラス名，レコード番号，文番号，開始位置，終了位置，単語，検出した辞書等の識別情報」で構成される CSV データ，すなわち，中間データを生成し，データベースに保存する．
(7) 中間データの解析状況を可視化して固有表現クラスの判定状況を確認する（例えば，図 2）と，スクリプトの実行直後は，人名クラスの判定箇所で，形態素解析で必要以上に単語が細分化されたり，区切りを誤ってい

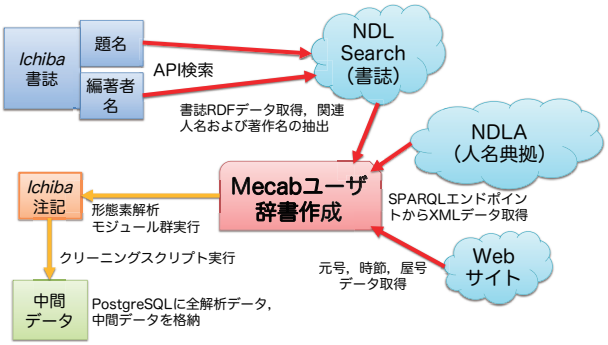

図 1　Mecab ユーザ辞書作成および解析スクリプトの動作．

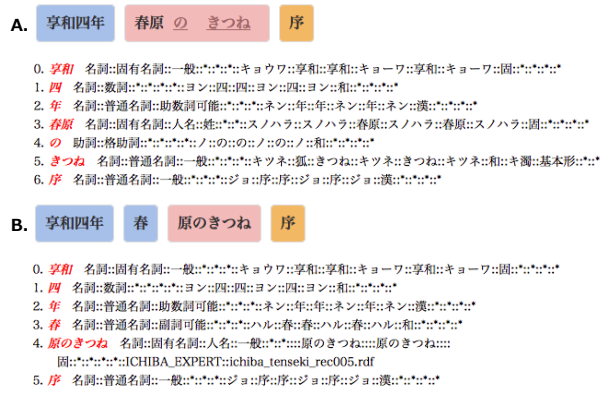

図 2 Mecab ユーザ辞書適用による固有表現抽出改善例．A は人名ユーザ辞書適用．B は人名ユーザ辞書適用後．基本辞書では「春原」（人名）と判定されるが，「原のきつね」が外部機関のデータを取り入れたユーザ辞書によって「人名」と判定されると，それに連動して「春」も時節の単語として適切に判定される．

る箇所がある．そのため，固有表現クラスの判定に重複や分断，誤判定が発生する．そこで，中間データに対し，一定の条件下でデータのクリーニングを行うようなスクリプトを追加で作成する．一定の条件とは，以下の条件を指す．

- 文中で人名クラスの単語が連続している，または一部重複している場合．
- 文の末尾等の不自然な箇所に時の固有表現がある場合．*ichiba* 注記文では，時を表す記述は文の先頭付近にある程度まとまって行われている．

(8) データベースに格納した中間データを基に，書誌データの構造化を行う．この部分は，OpenRefine[5] に RDF refine[24] モジュールを追加して，RDF データを生成する過程と同一である．最終的に構造化データを自動で生成する．

4. 結果

4.1 注記文における固有表現の抽出精度の測定方法

データベースに格納した単語データと，あらかじめ手動で作成した正解データと比較し，注記文における固有表現の抽出精度を求める．特に，人名および著作名（書名）は古典籍の研究で重要であるが，一般的な語ではないため，Mecab 基本辞書には未登録の単語が多い．そこで，Web に既出の人名を登録したユーザ辞書および人名推定モジュール（3.2.3 節 (4) 参照）の有無により，固有表現単語の抽出精度がどの程度推移するか調査する．

辞書セットおよびモジュールの組み合わせ一覧を表 4 に示す．Web で収集した人名の知識を基に作成したユーザ辞書群（表 4 のユーザ辞書 1）とその他の時節，部品，ロール，専門用語を登録した辞書群（表 4 のユーザ辞書 2）に分け，辞書の有無で固有表現の抽出精度を測定する．また，人名推定モジュール（3.2.3 節 (4) 参照）の適用有無についても測定する．その際，全ての組み合わせにおいて，UniDic の基本辞書は使用する．

表 4 辞書セットおよび人名推定モジュール（3.2.3 節 (4) 参照）の組み合わせ一覧．○は適用有り，×は適用無し．

組み合わせ	ユーザ辞書 1	ユーザ辞書 2	人名推定モジュール
A	×	×	×
B	×	×	○
C	○	×	×
D	×	○	×
E	○	×	○
F	×	○	○
G	○	○	×
H	○	○	○

ユーザ辞書 1: 時節，部品，ロール，専門用語を登録した辞書
ユーザ辞書 2: NDLA，NDL サーチ，*ichiba* 編著者名を登録した辞書

ただし，江戸時代の人名および書肆（しょし）（出版者）には地名が付加されている場合（江戸上総屋利兵衛，東京金港堂書籍株式会社等）が多い．そのため，地名と人名の識別が困難である．そこで，精度判定の際に，地名に人名および団体名が続く場合は，地名を含んだ人名であっても人名部分を全て検出している場合は正解とする．

4.2 注記文における固有表現の抽出精度の測定結果

辞書セットおよびモジュールの組み合わせ（表 4）による固有表現抽出スクリプト（3.2.3）の実行を行い，手動でカウントした正解データ（表 5）と照合して抽出精度を測定した．その結果，表 6 に示す結果を得た．また，表 6 を図 3 に示した．

表 5 手動でカウントした *ichiba* 注記文中の固有表現クラス別正解数．

固有表現クラス	正解数
時	244
著作名	591
人名	442
地名	85
部品名	376
ロール名	162

その結果，2 つの辞書群を全く使用しない場合（表 4 の A）よりも，使用した方（表 4 の G）が全体的な精度が上昇した．特に，人名の F 値は 0.217 から 0.426 へと倍となった．加えて，ユーザ辞書適用で，隣接する単語の固有表現も適切に抽出される箇所が増えた（図 2）．しかし，この結果は，固有表現の抽出を支援するという意味で実用的な水準（本研究では 0.900 以上とする．）の半分以下であった．

表 6 辞書セットおよびモジュールの組み合わせ別の固有表現抽出精度（F 値）.

種名	時	著作名	人名	地名	部品名	ロール名
A	0.936	0.965	0.217	0.521	0.928	0.901
B	0.936	0.965	0.825	0.580	0.928	0.901
C	0.945	0.965	0.217	0.521	1.000	0.932
D	0.975	0.965	0.426	0.661	1.000	0.994
E	0.945	0.965	0.857	0.580	1.000	0.932
F	0.967	0.965	0.878	0.779	0.929	0.952
G	0.975	0.965	0.426	0.661	1.000	0.994
H	0.975	0.965	0.914	0.779	1.000	0.994

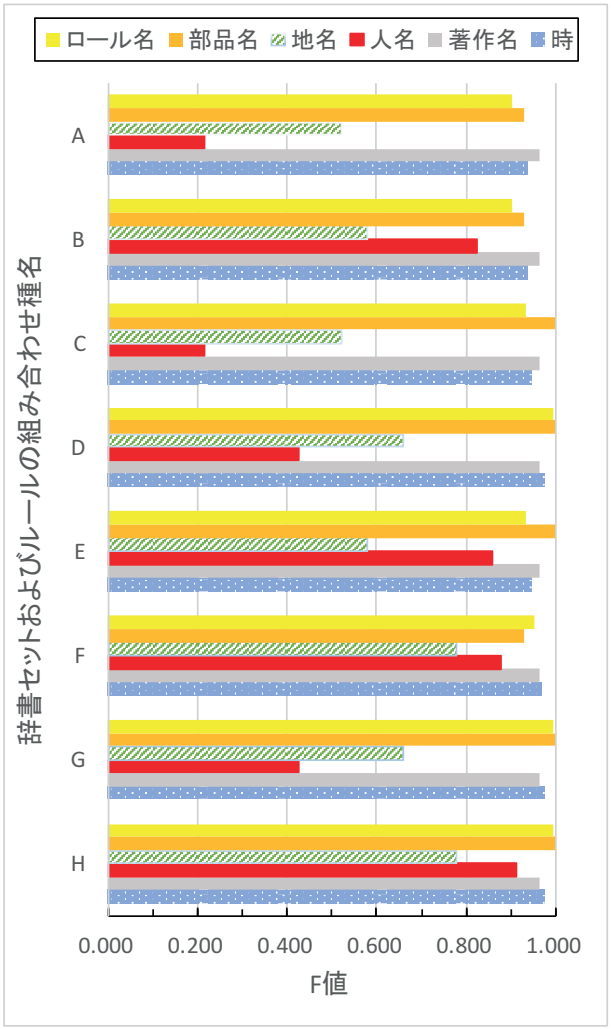

図 3 辞書セットおよびルールの組み合わせによる固有表現クラス別 F 値の推移.

しかし，人名推定スクリプトを Web データを集めた辞書抜きで行った結果（表 4 の B），F 値は 0.825 と 2 つの辞書群を用いて人名推定モジュールを用いない場合（表 4 の G）の F 値 0.426 と比較して倍近く精度が高まった．さらに，人名推定モジュールおよびユーザ辞書 1 および 2 を全て適用すると，人名の精度は 0.914 と実用的な水準に高まった．

人名および地名を除く他の固有表現クラスは，Mecab の基本辞書のみで 0.900 以上の精度を示した．また，ユーザ辞書を適用することで，時，人名，地名，部品名，ロール名の抽出精度がそれぞれ 0.936 から 0.975，0.217 から 0.426，0.521 から 0.661，0.928 から 1.000，0.901 から 0.994 に上昇した（表 4 の A および G）．地名は Mecab 基本辞書に登録された単語で形態素解析されているが，地名以外の単語の抽出精度が上昇すると，連動して精度が 0.521 から 0.779 に上昇した（表 4 の A および H）．

5. まとめ

形態素解析を用いて，注記文から古典籍の理解に必要な固有情報を半自動的に取り出すための方法を提案した．古典籍研究に昼用な固有表現と，オンラインで得られる古典籍に関する単語群をユーザ辞書として事前に用意した．また，辞書に登録されていない人名を効率よく推定するための推定規則を策定した．

注記文中には，Web で収集可能な既知の単語が少ないため，ユーザ辞書のみを適用するよりも，人名推定モジュールのような仕組みを用いて抽出するほうが精度が高いという結果を得た．この人名推定モジュールは，注記文中に部品名およびロール名がある場合に近隣の未知語は人名であると推定するというルールで動作する．すなわち，古典籍の注記文の場合，部品名およびロール名が人名を判定する鍵となることを示している．

固有表現の抽出精度を実用的な水準にするためには，人名周辺の，時，著作名，地名，部品名，ロール名に関する知識を利用することで，良好な結果を得られる．加えて，書誌が対象とする時代背景に関する知識をユーザ辞書とスクリプト中の固有表現判定ルールとして補うことも精度を高める有効な手段である．

今後は，固有表現抽出スクリプトに用いた注記文例数が少ないため，サンプル文例を増やし，上記の結果を改めて検証する予定である．また，古典籍以外の書誌にも部品名およびロール名を書誌記述の基点することが可能か検証を行う必要がある．

参考文献

[1] 国立国会図書館. 国立国会図書館デジタルコレクション. http://dl.ndl.go.jp/, 2016. 2016-01-01 参照.
[2] 国文学研究資料館. 日本古典籍総合目録データベース. http://base1.nijl.ac.jp/~tkoten/about.html, 2006. 2016-01-01 参照.
[3] 吉賀夏子, 渡辺健次, 只木進一. 貴重書メタデータの設計図としての書誌オントロジーを適用した linked data. 第 35 回セマンティックウェブとオントロジー研究会, No. SIG-SWO-035-08, pp. 1–9, 3 2015.
[4] 吉賀夏子, 渡辺健次, 只木進一. 貴重書中の部品を記述できるオントロジーに基づき linked data 化したメタデータを用いた人名ネットワーク構築の試み. 人工知能学会全国大会論文集, Vol. 29, No. 1G4-1in, pp. 1–4, 2015.

[5] OpenRefine Community. Openrefine. http://openrefine.org, 2012. 2016-01-01 参照.
[6] 井上敏幸（編）. 市場直次郎コレクション目録. 佐賀大学附属図書館・地域学歴史文化研究センター, 2007.
[7] 佐賀大学附属図書館. 佐賀大学貴重書コレクション. http://www.cc.saga-u.ac.jp/OgiNabesima, 2001. 2016-01-01 参照.
[8] 吉賀夏子, 渡辺健次, 只木進一. 書誌学的情報およびデータ入力を考慮した貴重書書誌オントロジーの構築. 第 28 回人工知能学会全国大会 (JSAI2014), No. 1G5-OS-19b-1, 5 2014. 2016-01-01 参照.
[9] 国立国会図書館. Web ndl authorities. http://id.ndl.go.jp/auth/ndla, 2011. 2014-04-02 参照.
[10] 加藤文彦. Linked data 作成支援ツールの現状と課題. 第 24 回セマンティックウェブとオントロジー研究会, No. SIG-SWO-A1101-03, pp. 1–4, 2011.
[11] 乾孝司, 村上浩司, 橋本泰一, 内海和夫, 石川正道. 接尾辞情報を利用した文書からの組織名抽出. 人工知能学会論文誌, Vol. 24, No. 6, pp. 469–479, 2009.
[12] Yaser Al-Onaizan and Kevin Knight. Translating named entities using monolingual and bilingual resources. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 400–408. Association for Computational Linguistics, 2002.
[13] Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff. Definition of the cidoc conceptual reference model. http://www.cidoc-crm.org/docs_crm_version_5.0.4.pdf, 2011. 2016-01-01 参照.
[14] Antoine Isaac. Edm primer. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf, July 2013. 2016-01-01 参照.
[15] 村田良二. ミュージアム資料情報構造化モデルの開発. 人文科学とコンピュータシンポジウム, pp. 63–70, 12 2005.
[16] 工藤拓. Mecab: Yet another part-of-speech and morphological analyzer. http://mecab.sourceforge.net, 2013. バージョン 0.996, 2016-01-01 参照.
[17] 国立国語研究所. Unidic. http://pj.ninjal.ac.jp/corpus_center/unidic/, 2013. バージョン 2.1.2, 2016-01-01 参照.
[18] 国立天文台. こよみ用語解説. http://eco.mtk.nao.ac.jp/koyomi/faq/24sekki.html, 1994. 2016-01-01 参照.
[19] みんなの知識委員会. みんなの知識ちょっと便利帳. http://www.benricho.org/koyomi/tuki.html, 2002. 2016-01-01 参照.
[20] ウィキペディア日本語版. 干支. https://ja.wikipedia.org/wiki/干支. 2015 年 12 月 30 日 (水) 11:04 の版, 2016-01-01 参照.
[21] ウィキペディア日本語版. 元号一覧（日本）. https://ja.wikipedia.org/wiki/元号一覧_(日本). 2015 年 12 月 14 日 (月) 09:46 の版，2016-01-01 参照.
[22] 国文学研究資料館. 利用のしかた（日本古典籍総合目録データベース）. http://base1.nijl.ac.jp/ tkoten/howto.html, 12 2006. 2016-01-01 参照.
[23] ウィキペディア日本語版. 屋号. https://ja.wikipedia.org/wiki/屋号. 2015 年 12 月 11 日 (金) 09:45 の版, 2016-01-01 参照.
[24] Fadi Maali and Richard Cyganiak. Rdf refine. http://refine.deri.ie, 2011. 2016-01-01 参照.

正誤表

第 109 回 人文科学とコンピュータ研究会発表原稿「古典籍書誌注記文からの作品構造および関連人物のつながりを明らかにする周辺情報抽出」について，下記の通り訂正します.

1. 4 ページ　表 3　部品名に対する単語の例
（誤）表示
（正）表紙

2. 5 ページ　4.1 節　第 2 段落 3 行目
（誤）表 4 のユーザ辞書 1
（正）表 4 のユーザ辞書 2

3. 5 ページ　4.1 節　第 2 段落 4 行目
（誤）表 4 のユーザ辞書 2
（正）表 4 のユーザ辞書 1

4. 6 ページ　5 節　第 1 段落 3 行目
（誤）昼用な
（正）重要な