

古典籍データ中の非正規記述からの周辺情報抽出とその構造化

吉賀 夏子・只木 進一（佐賀大学大学院 工学系研究科）

近世以前の書籍である古典籍の書誌情報の多くは、正規化されていない。特に、自然言語で記載されている注記などの記述は、人名、時、場所、関連作品などの周辺情報を含んでおり、対象書籍の成立過程について探索する鍵となり得る。古典籍の書誌情報が **Linked Data** として公開されることで、古典籍間の関係の研究に、一層利用されることが期待される。そこで、我々は、周辺情報の手がかりとなる固有表現を書誌のテキストから抽出することで、**Linked Data** に適したデータへ変換する手法を提案する。変換した書誌データは簡易的なオントロジーに従い、典拠情報と対応付けられる。

Extracting and Constructing Contextual Information from Unstructured Texts in Bibliographical Data of Japanese Historical Rare Books

Natsuko Yoshiga / Shin-ichi Tadaki
(Graduate School of Science and Engineering, Saga University)

Most of records in bibliographies for Japanese historical rare books are not normalized. Especially, notes, which are written in natural language, contain contextual information such as human names, dates, and places which may be keys to explore history of a book. Those bibliographical data will be utilized for exploring relations between historical books, if they are published as a linked data. For this purpose, by extracting proper expressions included in notes containing key contextual information, we propose a method to convert those texts into well-formed data suitable for linked data. Those data will obey a simple ontology and be provided with a URI to an authorized data entity.

1. はじめに

近世以前の書籍である古典籍の書誌情報が様々な規模の機関によって Web 公開されている。この情報の中に含まれる、作品に関わる人物、時代、地理、関連作品といった周辺情報は、研究者が求める重要な情報である[1]。しかし、これらの情報の多くは、人が読んで理解することを前提とした、正規化されていないデータである。そのため、書誌をまたがって項目間の関連を自動的に取得することはできない。その結果、研究者は、書籍あるいは著者などの人の関係を得るために、書誌情報と書籍を手探りで読み解かなければならない。

周辺情報に含まれる人物や場所などの基本情報の多くは Web で参照できる。既に、国立国会図書館などは基本情報を構造化データ (**Linked Data**[2]) として公開している。これら構造化された基本情報を古典籍などの文化財書誌の検索に活用するには多くの困難がある。その原因は、元の古典籍書誌の属性と値が、外部の **Linked Data** に結びつくように正規化されていないためである。正規化を実際に行うには、書誌データの記載

内容を把握できる専門知識を基に、属性と値に相当する語を抽出し、適切な組にする必要がある。しかし、膨大に存在する書誌データを人手で正規化するのは困難である。

本研究では、書誌情報の注記に含まれる作品の周辺情報までもデータ化し、かつ機械可読とするために、自然言語で記述される作品を構成する部品に関連する場所、時、人物などの属性及び構造を取得する手法を、具体例に基づいて提案する。提案手法は、最終的な確認は人手で行うことを前提として、主に書誌中の重要な周辺情報を抽出し、正規化する作業を大幅に軽減する。

これらの属性をデータ化し、相互に関連づけることで、書誌データの内外を問わず、作品の周辺情報を構造として取得することが可能となる。

これまで、川村と大須賀[3]、Thompson Reuter社[4]をはじめとした研究で、非構造化テキストから **Linked Data** への変換手法の提案が行なわれている。本研究では、これらの手法を拡張し、古典籍書誌データ中に注記として記載された非構造化テキストなどを、自然言語処理を活用して属性と値の組（以下、「トリプル」と呼ぶ）を抽出し、構造化データとする。

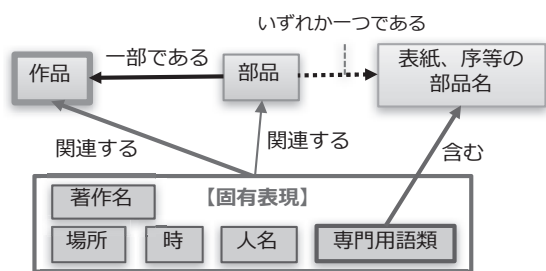


図1 部品の集合体としての作品と部品、部品と固有表現の関係。

提案手法では、はじめに、注記を含む書誌データ全体から、時、場所、人名、部品名、ロール名（作、著、画などの役割を示す語句）、古典籍の書誌で使用されている専門用語、著作名をトリプルとして抽出する。以下、これらを「固有表現」と呼ぶ。

次に、抽出したトリプルが作品本体とどのような関係を持つのかをトリプルとし記述するために、古典籍の記述に必要な属性の定義と属性同士の関係についての情報もデータ化する。そのために、図1に示すように、作品を書誌記述の対象となる項目（表紙、見返しなどの部位を表す語彙）をラベルとした部品と捉え、部品とトリプルの関係を定義する。

2. 非構造化テキスト文を含む書誌データからのトリプル抽出およびURI付与

2.1 古典籍書誌データをLinked Data化する上での問題

古典籍書誌データの多くは、古典籍を直に調査

表1 *ichiba* のレコード記述例。赤字は部品およびロール名。

属性番号	属性	値
0	整理番号	45
1	分類	5-1-(2)-(6)-(6)
2	分類1	文学
3	分類2	国文
4	分類3	小説
5	分類4	(データなし)
6	分類5	洒落本
7	順位	(データなし)
8	画像公開	(データなし)
9	書名	青楼心得唄
10	読み	セイロウココロエグサ
11	書型	(データなし)
12	巻冊	(データなし)
13	編著者	(二世) 蓬萊山人作
14	刊行・書写年次	安政四年序
15	西暦	1857
16	刊写	刊
17	注記	安政四年一月蓬萊山人序。天明五年刊「息子部屋」ノ改題本ナリ。
18	印記	1
19	文庫	市場
20	他	(データなし)

する研究者によって手書きされた調査票を基に作成される（表1）。

しかし、現代書籍に比して、古典籍では、書籍名、著者名、出版名などが、判読困難あるいは無記載である場合があり、データ登録者の主観によるデータの揺れが混入しやすい。そのため、データが十分に正規化されていない場合が多い[5]。さらに、書籍を構成する部品（序や書籍を構成する小作品）の著者や時期に関する情報が、書籍の成立や関連を研究する研究者にとって重要であるにも関わらず、書誌項目のなかで自然言語によって記述されている。例えば、表1において、編著者、刊行・書写年次および注記中の「序、作、刊」という文字が、人名や時を表す語に接続されている。その結果として、それらの重要な情報が、非構造的・非正規化情報としてしか与えられないという大きな問題がある。

そのため、仮に、表1に挙げる各属性に Dublin Core [6]や DBPedia Japanese[7]などのメタデータを直接対応させてトリプルを抽出したとしても、古典籍研究者が求める周辺情報を取得可能な粒度で得ることは困難である。

2.2 周辺情報獲得を目的とした非構造化データからの固有表現抽出に関する既往研究

そこで、古典籍書誌からトリプルを抽出するには、正規化されていない属性、特に注記に記述された情報から、自然言語処理で周辺情報の手がかりとなる人名、時などの固有表現を抜き出す。

一般的に、抽出する固有表現は対象分野によって異なるが、基本的に、時、場所、人名および組織名、その他対象分野に関連深い固有名詞、名詞同士の関係を示す述語および補語が採用される。

人文系の書誌データを対象にした自然言語処理については、似内と奥野[8]が歴史コンテンツに含まれる非構造化データ中の、時、出来事、関連人物などに着目し、固有表現抽出を試みている。

また、我々も[1]において、古典籍分野の用語を収録したユーザ辞書を搭載する形態素解析ツール Mecab[9]と簡易的な文脈判定を用いて、書誌データのテキストから固有表現を抽出している。

加えて、震災時の行動分析やソーシャルメディアとマスメディアの意見分析の視覚化を目的とした研究群[3,10]においても、形態素解析と係り受け構文解析が基礎技術として採用されている。

これらの研究では、有益な周辺情報が構造化されていないテキストに含まれていることに着目し、それらの情報を Linked Data として抽出することで活用可能性を拡大することを目的としている。

2.3 トリプル抽出手順

本研究では、佐賀大学附属図書館所蔵の市場直次郎コレクション[11]（以下、*ichiba* と呼ぶ）を例に、前述の研究で行なわれている自然言語処理を応用し、トリプル抽出と正規化を行う。

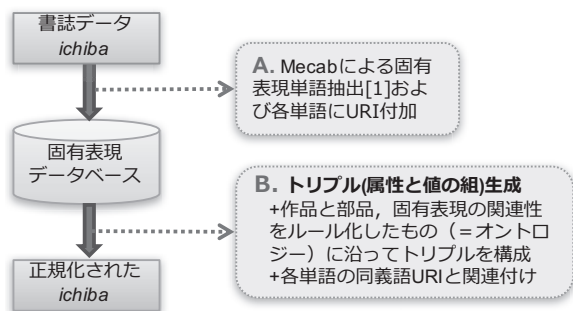


図2 書誌データ *ichiba* からのトリプル抽出手順.

古典籍研究, 特に古典籍間の関係や来歴に関する研究では, 書籍の部品毎の著者や時期に関する情報は重要である. そのため, 古典籍書誌の目録記述規則や凡例においても, 作品の識別に必要な要素として, 表紙, 序など, 作品の部品に関する記述が言及されている[12,13,14]. しかし, *ichiba* を含む多くの場合は, これらの情報が注記などに自然言語で記述されている.

そこで, これら部品の来歴に関する情報を注記などから抽出し, 古典籍に関する専門知識を参照しつつ, 構造化することを目指す. 構造化したデータには, 典籍データへの URI を付与する. 処理の全体的な流れを図2に示す.

2.4 単語抽出および固有表現判定

はじめに, 古典籍分野の専門用語を収集した辞書を準備する, この辞書をユーザ辞書として, 形態素解析ツール Mecab[9]と簡易な文脈判定スクリプトを用いて, 非構造化テキストを分析する.

事前に準備する Mecab 用ユーザ辞書には, 固有表現として設定した時, 場所, 人名, 部品名, ロール名, 専門用語, 著作名に関連する時節用語, 十干十二支, 年号, 人名典拠の標目および別名データ, 目録記述規則や *ichiba* 凡例で使用されている専門用語類を抽出して登録した(詳細および出典は[1]の表3参照). これらの単語は, 古典籍の書誌を読むために必要な予備知識として位置付けられる.

ichiba 中の 211 レコードに含まれる 951 の注釈文に対して Mecab による単語への分解を実施した. 正解の判定は Mecab 付属の評価ツール及び別途準備した 4449 単語のデータを用いた. 単語への分解精度の F 値は 0.863 であり, ユーザ辞書を用いない場合の値 0.594 と比して大幅に改善した.

次に, 分割された全ての単語に対して, 固有表現か否かを判定する. 固有表現と判定された単語には固有表現種別, レコード番号, *ichiba* 属性番号(表1参照), 文番号(注記内の文に割り当てた番号), 単語区切り開始位置, 終了位置, 単語をタグ付けしてデータベースに登録する.

ユーザ辞書中にない人名は, 注記記載の法則性から推論することがある程度可能である. 図3で

例1) 嘉永四年立秋廓遊舎文盛序。

／はMecabによる単語区切り位置。

1	Mecab単語分割	嘉永／四／年／立秋／廓／遊／舎／文／盛／序／。
2	日付・専門用語類の同定	嘉永四年／立秋／廓／遊／舎／文／盛／序
3	部品・ロール名が文末にある文章に対し, 人名推定	嘉永四年／立秋／廓／遊／舎／文／盛／序 (E) ← (S)
4	人名ラベル付与	嘉永四年／立秋／廓遊舎文盛／序
5	クリーニング	嘉永四年立秋／廓遊舎文盛／序

*パターンマッチ 条件1: 「の」以外の助詞ではない。
条件2 「の」が単語の両端にある時は人名にせず探索続行。
条件3: 記号ではない。
条件4: 他の固有表現ではない。

例2) 大坂和多屋平兵衛

／はMecabによる単語区切り位置。

1	Mecab単語分割	大坂／和／多屋／平兵衛
2	地名・Mecabによる人名の同定	大坂／和／多屋／平兵衛
3	屋号・称号が文中にある文章に対し, 人名推定	大坂／和／多屋平兵衛 (E) ← (S)
4	人名ラベル付与	大坂／和多屋／平兵衛
5	クリーニング	大坂／和多屋平兵衛

**屋号: 屋, 堂, 亭, 軒, 家, 社, 庵など
***称号: 翁, 氏, 散人, 山人, 産人, 主人など

図3 人名推定モジュールの仕組みおよび適用例.

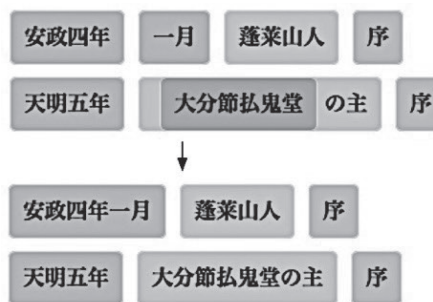


図4 クリーニングスクリプト適用例. 上: 適用前, 下: 適用後. 各単語のラベル色は, 青ラベル: 時, 赤: 人名, 橙: 部品名の固有表現を表す.

は, *ichiba* 注記の文章パターンの一つで, 部品名が最後に位置する場合と, 屋号[15]および称号が文中に出現する場合の経験的推論則による人名推定手順を説明している.

最後に, 固有表現をデータベースに登録後, 固有表現付与の結果には, 日付の判定や先述の推定モジュールの影響で人名の文脈判定の結果, 重複や余分な分割が見られる. そのため, 重複や分割を修正するクリーニングスクリプトを適用する(図3および図4).

2.5 分割単語に対する URI 付与

次に, 前述の固有表現データベースに登録した単語に一意の URI を振る.

Linked Data の作成において, 基本的にトリプルの構成要素は URI で表現する. そのため, 抽出した全単語に独自の URI を付与する. その際, 固有表現につながる単語が抽出元のテキストのどの部分から抽出されたものを明示する. 典拠を示すことで人手による修正も容易となる.

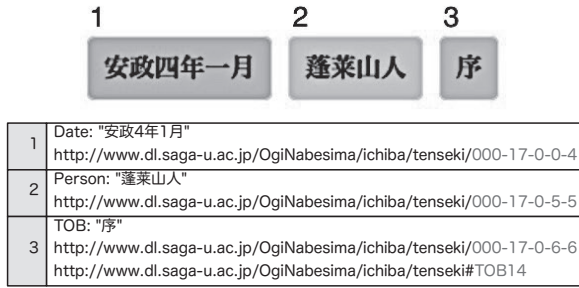


図5 自然言語処理により抽出した単語に対して固有表現およびURIを付与した結果。固有表現ラベルであるDateは時、Personは人名、TOB (Terminology of bookbinding) は部品名を示す。

同様の手法はThompson Reuter社のOpen Calaisで行われている[4]。Open Calaisでは、ニュース記事などの非構造化テキストから企業、人名、地名、取引や活動の分類名などを表す単語を自動抽出する。抽出単語には、典拠および元テキスト中に含まれる他の単語との共起および位置関係から算出した信頼度と独自URIが付与される。各単語同士の関係はOpen Calais Ontologyに基づき構造化される。

本研究では、Open Calaisの手法を応用し、固有表現と判定された単語に対して自動でURIを付与する。その具体例を図5に示す。自動付与された各URI末尾の赤字(“000-17-0-0-4”など)はレコード番号、属性番号(17は注記)、文番号(注記内の文に割り当てた番号)、単語区切り開始位置、終了位置を示す。部品名、ロール名、専門用語には、別途クラス定義のためクラス名および番号を付与した。

この方式は、自動抽出した単語がURIに適さない記号や文字列であっても、解析元での単語の位置をURIとして表現可能である。同時に、人にとっても抽出単語の出所の把握が容易である。

3. 作品構成部品の特定とその構造化

3.1 目録記述規則を基にした作品構成部品の

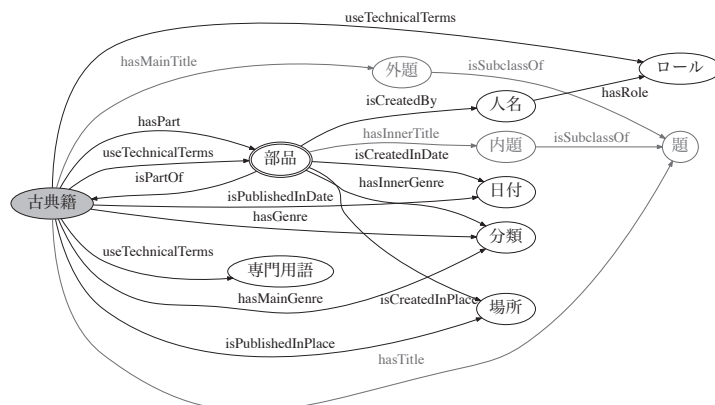


図6 古典籍書誌に存在する概念クラス(部品および固有表現)の関係。右表は左図中の概念クラス間の関係ラベル名および定義。

hasPart	A hasPart Bで古典籍Aは部品Bを持つ。
hasRole	A hasRole Bで人名Aは役割Bを持つ。
useTechnicalTerms	A useTechnicalTerms Bで古典籍Aは専門用語Bを持つ。
isPublishedInPlace	A isPublishedInPlace Bで古典籍Aは場所Bで出版される。
isPublishedBy	A isPublishedBy Bで古典籍Aは人名Bにより出版される。
isPublishedInDate	A isPublishedInDate Bで古典籍Aは日付Bに出版される。
hasTitle	A hasTitle Bで古典籍Aは題Bを持つ。
hasInnerTitle	A hasInnerTitle Bで部品Aは内題Bを持つ。
hasMainTitle	A hasMainTitle Bで古典籍AはBの外題を持つ。
hasGenre	A hasGenre Bで古典籍Aは主分類名Bを持つ。
hasMainGenre	A hasMainGenre Bで古典籍Aは分類名Bを持つ。
hasInnerGenre	A hasInnerGenre Bで部品Aは分類名Bを持つ。
isPartOf	A isPartOf BでAは古典籍Bの部品である。
isCreatedBy	A isCreatedBy Bで部品Aは人名Bによって作成される。
isCreatedInPlace	A isCreatedInPlace Bで部品Aは場所Bで作成される。
isCreatedInDate	A isCreatedInDate Bで部品Aは日付Bで作成される。

特定

研究者にとっては古典籍の記載内容のみならず、書籍そのものの作りや書き込み、奥付なども有益な情報である。そのため、作品の内容そのもののみでなく、表紙、見返しなど、作品を構成する部品についても、書誌記述の対象となる[16]。

図1で示した作品と部品の関係について、固有表現と合わせて詳細化したものを図6に示す。

本研究での部品とは、書誌中で物理あるいは論理的な作品の一部として特定できる箇所とする。具体的には、表紙画、挿絵、序跋などである。これらの部品には、本体とは異なる時、場所、人が関わっている場合がある。

部品についてのこうした記述に対し、作品全体の書誌と同様の粒度で記載する仕組みを適用すれば、作品の成立につながる周辺情報をより詳細に掘むことが可能となる。

そこで、本節では、2節で抽出した部品名、ロール名および専門用語に相当する単語を、あらかじめ書誌学用語集、古典籍書誌記述目録など[11,12,13,14]を基に決定し、それらの記述規則を機械可読にする。Ichiba中に出現した部品名の数は表記ゆれも含めて86、ロール名は57、部品およびロール名以外の古典籍書誌で用いられる専門用語は84である。

これらの用語類は、ichibaのみならず国文学研究資料館の公開古典籍書誌[17]、国立国会図書館デジタルコレクション内書誌(詳細レコード)[18]などで使用されている。例えば、部品「序」は、ichibaでは、全レコード中49.3%、[17]では30.4%の頻度で使用されている。他の古典籍書誌が[12,13,14]などの記述規則に準じている場合、部品などの用語群を活用する方法は有効である。

3.2 古典籍中の部品および固有表現の関係構造のデータ化

また、部品名などの用語群には、その定義のみならず、表記ゆれ、意味の重複、包含関係、複合語など、詳細な関係がある。加えて、古典籍書誌でのみ言及される、固有表現と部品間の特別な関

係がある。

例えば、固有表現の一つである題には、外題と内題と呼ばれる関係がある（図6赤文字部分）。古典籍のような古い書物は表紙に題が明確に記載されていない場合が多いためである。そこで、便宜的に見返しや本文の最初の一文などを取って題とする場合がある。この時の書籍内側の題を内題、外側の表紙や題箋（題を書き表紙に貼り付けるための付箋）の題を外題と呼ぶ。

内題と外題のような関係は、古典籍書誌を読み解く知識として、通常は書誌学に詳しい研究者が主に共有する。

そこで、図6の固有表現、部品などの書誌に出現する値をグループ化したクラスおよびクラスが内包する値（個体）について、定義域および値域、関係性などを設定したルールを記述し、機械可読にすることで、あらかじめ設定された属性（表1）に加えて書誌学的なルールに則ったデータの探索が可能になる。

具体的には、図6の中で、3つの用語クラス（部品名、ロール名および専門用語）が内包する各用語の関係について、表2に示すオントロジー表現データと呼ぶ構造を便宜的に作成する。

このオントロジー表現を表す式は、オントロジー言語 OWL[19]で使用するクラス記述に対応させる。OWLは対象の知識（ここでは古典籍書誌についての知識を指す。）を機械可読にするためにW3Cが標準化した言語である。

もちろん、Protégé[20]などのオントロジーツールのみを用いてクラス、書誌属性、値そのもの（個体）間にまたがる関係を構築することは可能である。しかし、それらの数が多い場合は、ルール構築作業がかえって煩雑になる。そこで、古典籍の用語は同義関係あるいは包含関係などの基本的なルール構築をオントロジー表現で行っておき、スクリプトでRDFおよびOWL形式データに自動変換する。

表2 オントロジー表現データ。左辺の語について、=は右辺の語と同義である、:は右辺の語を属性として持つ、⊂は右辺の語に内包されるという意味。

値番号	値	オントロジー表現	対応スキーマでの表現
TOB00	内題	(なし)	rdf:type TOB
TOB14	序	(なし)	rdf:type TOB
TOB15	〈序〉	=序	owl:sameAs TOB14
TOB16	(序)	=序	owl:sameAs TOB14
TOB17	序文	=序	owl:sameAs TOB14
TOB18	自序	=(序:著者)	rdf:type TOB14 ichibaTenseki:isCreatedBy
TOB19	序中	=序	owl:sameAs TOB14
TOB20	序題	=(序:題) ⊂内題	rdf:type TOB14 ichibaTenseki:hasInnerTitle および rdfs:subClassOf TOB00
TOB21	序中題	=序題	owl:sameAs TOB20

接頭辞	名前空間
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
ichibaTenseki	http://www.dl.saga-u.ac.jp/OgiNabesima/ichiba/tenseki#

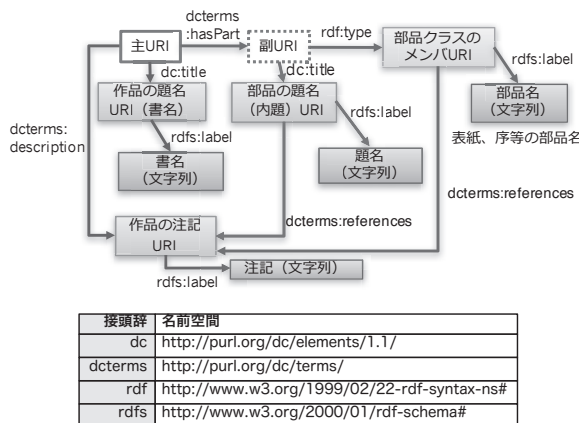


図7 主および副 URI が紐づいた古典籍作品および部品、題名、自動抽出した単語の典拠が注記である場合の関係。下表はノード間の関係に用いた標準スキーマ。

その際、作品本体に主 URI、部品に副 URI の設定を行う（図7）。副 URI にアクセスすることで、書誌データから抽出したトリプルがどの部品に由来するかが機械可読となる。

3.3 書誌データと外部典拠情報の関連付け

加えて、件名および人名典拠データを持つNDLA[21]などの Linked Data 上も自動探索し、対象の単語（値）が同種の固有表現である場合、その候補は同義語であるとみなし、トリプルに組み込む。ただし、同義語の候補が複数出現する場合、手作業で選択する。

さらに、各部品の名称の同義語を、The Art &

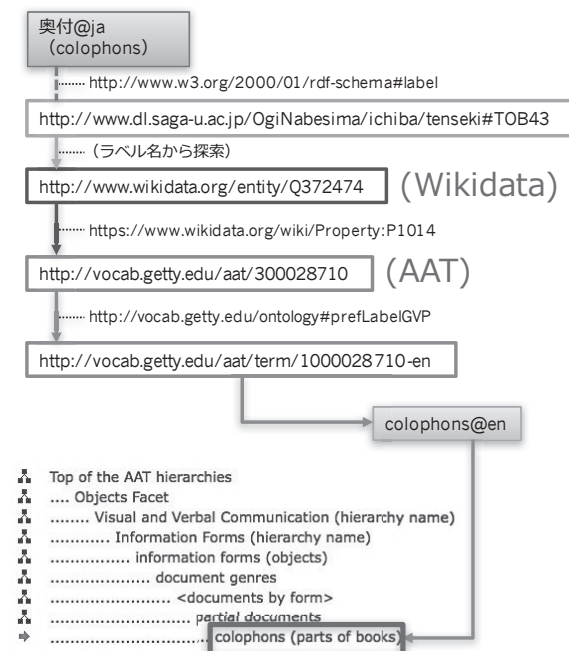


図8 部品から典拠データを探す例。Wikidata[22]経由で奥付 (colophons) に対する AAT 内の同義語 URI を探索。

Architecture Thesaurus® (AAT) [23]や、NDLA[21]など、外部組織の部品名 URI および定義を持つ Linked Data から自動探索する (図 8)。

設定した部品名と外部組織で定義された典拠件名が探索経路から同義だと判明すれば、その典拠件名に紐づく外部資料も検索の対象になる。

4. まとめ

本研究では、手間の掛かる書誌データ正規化および固有表現の抽出作業を可能な限り軽減するため、固有表現が含まれる単語を自然言語処理技術で抽出し、URI を付与する仕組み (2 節) および書誌記述対象に沿った作品の部品化と構造のデータ化 (3 節) を提案した。

古典籍書誌 *ichiba* に対し、提案手法を適用したところ、作品のどの部分の由来なのかを含めて正規化されたトリプルが生成された。トリプル中の属性および値には URI が付加されているため、書誌中の記述を URI で指定可能となった。

正規化された書誌データに対して特定の人名の検索を行うと、作品成立に貢献した部位、時、場所、人間関係などの、元書誌が持つ情報量を超えた周辺情報も合わせて自動的に取得できる [10,16]。

今後は、2 節での自然言語処理で、単語の切り出しと固有表現の推定精度を高める必要がある。3 節では、部品および個体に対応する、外部組織の Linked Data を自動探索する技術および手順を改良する必要がある。

また、古典籍書誌の固有表現の抽出と構造化を、ウェブ公開されている古典籍書誌のみならず、古文書など別の文化財の公開書誌データに適用可能か検証を行う予定である。

参考文献

- 1) 吉賀夏子, 渡辺健次, 只木進一: 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, 情報処理学会研究報告, 2016-CH-109, No. 3, pp. 1-7 (2016).
- 2) 武田英明 (監訳): *Linked Data*, 近代科学社, (2013).
- 3) 川村隆浩, 大須賀昭彦: TEXT2LOD:~テキスト情報の LOD 化に向けた Web API の開発~, 人工知能学会論文誌, 1346-0714, <http://dx.doi.org/10.1527/tjsai.LOD-21> (参照 2016-10-28), 一般社団法人人工知能学会, (2016).
- 4) Ulicny, Brian: *Constructing Knowledge Graphs with Trust*, 4th International Workshop on Methods for Establishing Trust of (Open) Data, Bentlehem, USA, (2015).
- 5) 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, No. 45, pp. 3-9 (2013).
- 6) DCMI: *Dublin Core Metadata Initiative (DCMI)*, <http://dublincore.org> (参照 2016-10-25).

- 7) DBpedia Community: *DBpedia Japanese*, <http://ja.dbpedia.org> (参照 2016-10-25).
- 8) 似内勇太, 奥野拓: 固有表現抽出を用いた歴史コンテンツの固有表現抽出を用いた歴史コンテンツの LOD 化支援, *じんもんこん 2014 論文集*, 3, pp. 193-198, (2014).
- 9) 工藤拓: *Mecab: Yet another part-of-speech and morphological analyzer*, <http://taku910.github.io/mecab/>, ver. 0.996, (参照 2016-10-31).
- 10) 田代和浩, et al.: *Linked Data を用いたソーシャルメディア×マスメディアの比較実験*, 人工知能学会全国大会論文集, 2N1-OS-10d-3, 27, pp. 1-4, (2013).
- 11) 佐賀大学附属図書館: *佐賀大学貴重書コレクション*, <http://www.dl.saga-u.ac.jp/OgiNabesima> (参照 2016-10-28).
- 12) 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> (参照 2016-10-25).
- 13) 国立国会図書館: *日本目録規則 1987 年版改訂 3 版 和古書適用細則*, <http://warp.da.ndl.go.jp/info:ndljp/pid/9484238/www.ndl.go.jp/jp/library/data/wakosho201201.pdf> (参照 2016-10-28), (2012).
- 14) Isamu Tsuchitani, Manae Fujishiro: *DESCRIPTIVE CATALOGING GUIDELINES FOR PRE-MEIJI JAPANESE BOOKS*, Library of Congress, http://www.eastasianlib.org/cjm/jrb/2011_japanese_RareBooksCatalogingGuidelines.pdf (参照 2016-10-31), (2011).
- 15) ウィキペディア日本語版: *屋号 (日本についての項)*, <https://ja.wikipedia.org/wiki/屋号> (参照 2016-10-20), (2015).
- 16) 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).
- 17) 国文学研究資料館: *国文研古典籍データセット (第 0.1 版) 全件文書誌データ*, http://jcbv.nii.ac.jp/oa/NIJL0-1/NIJL0-1_Biblio-2.txt (参照 2016-10-30), (2012).
- 18) 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/> (参照 2016-10-30), (2011).
- 19) W3C: *OWL 2 Web Ontology Language Primer (Second Edition)*, <https://www.w3.org/TR/owl2-primer/> (参照 2016-10-20), (2012).
- 20) Stanford University: *Protégé*, <http://protege.stanford.edu> (参照 2016-10-25).
- 21) 国立国会図書館: *Web NDL Authorities*, <http://id.ndl.go.jp/auth/ndla> (参照 2016-10-28).
- 22) Wikimedia Foundation, Inc.: *Wikidata*, <https://www.wikidata.org/> (参照 2016-10-31).
- 23) J. Paul Getty Trust: *The Getty Vocabularies*, <http://vocab.getty.edu> (参照 2016-10-31).