

# Linked Data を用いた軍艦オントロジーの構築と平賀譲デジタルアーカイブへの活用

中村 覚 (東京大学 情報基盤センター)

本研究では、大日本帝国海軍造船中将であった平賀譲が遺した艦艇に関する設計図面や技術文書を公開する平賀譲デジタルアーカイブの利活用に向け、軍艦に関する知識の体系化を行うオントロジーの構築を行う。DBpedia を用いたオントロジー構築を行うことで、既存の知識ベースが提供するデータの再利用が可能となり、123件の軍艦から構成されるオントロジーを低コストで構築することができた。また、構築した軍艦オントロジーを用いて、軍艦の建造史や階層関係等を可視化することにより、軍艦に関する多角的な情報提供と、平賀譲デジタルアーカイブが公開する資料に対する多角的なアクセスを実現した。

## Development of Warship Ontology and Application to Yuzuru Hiraga Digital Archive with Linked Data

Satoru Nakamura (Information Technology Center, the University of Tokyo)

This study builds an ontology for systematizing knowledge about warships towards utilization of Yuzuru Hiraga Digital Archive, which publishes design drawings and technical documents relating to warships of the Japan Imperial Navy. Constructing an ontology using DBpedia enabled to reuse the data provided by existing knowledge bases and construct an ontology consisting of 123 warships at low cost. In addition, visualizing the history of construction and the hierarchical relationship of warships by use of the constructed ontology enables to provide multilateral information on warships and multifaceted access to the documents of Yuzuru Hiraga Digital Archive.

### 1. 緒言

東京大学が公開しているデジタルアーカイブの一つに、平賀譲デジタルアーカイブ[1]がある。本デジタルアーカイブは大日本帝国海軍造船中将および東京帝国大学第13代総長であった平賀譲が遺した資料をデジタル化し、インターネット上で公開するシステムである。この資料の特徴の一つとして、平賀譲が造船官として設計等に携わった軍艦に関する設計図面や技術文書を数多く含む点が挙げられる。特に図面資料等については、国内外を問わず軍艦史や科学史の研究において有益な資料である。

このような資料を公開するシステムにおいて、本研究では「多言語対応」「知識体系の整理」の2点の課題を取り上げる。先述したように、平賀譲デジタルアーカイブ(以下、平賀譲DA)には軍艦の設計図面や技術文書が数多く含まれる。一方、現在のシステムは日本語表記の目録データを用いた資料の検索や閲覧機能の提供にとどまる。そのため、日本語に馴染みのないユーザによる資料へのアクセスビリティは著しく低い点が課題の一つである。2点目の課題は、対象資料群に関する知識体系が整理されていない点である。特に、軍艦をはじめとする専門領域に関する資料を取り扱うため、資料の理解・把握には、戦

艦や航空母艦等の艦種、「長門型」や「大和型」等の艦型等に関する知識が求められる。対象分野に馴染みのない利用者に対する情報提供や、多角的な資料へのアクセスに向け、これらの知識体系の整備が必要である。

1点目の課題を解決する方法として、機械翻訳や人手による目録データの多言語化等が考えられるが、本研究では知識体系の整備の必要性を加味し、オントロジーの構築による課題解決を目指す。オントロジーとは、対象とする領域や組織、タスクに関する知識を体系化し、計算機による知識処理を可能とする技術である[2]。知識の体系化による資料へのアクセスビリティの向上を目的とした研究として、研谷[3]による人名典拠情報の構築や、安田ら[4]による同人創作物に関するメタデータモデルの構築等が挙げられる。また、Wikipedia等の知識ベースを活用したオントロジー構築に関する研究も行われている[5]。玉川[6]らは日本語WikipediaのリダイレクトリンクやInfoboxテンプレート等の様々なリソースを利用し、大規模かつ汎用的なオントロジー「日本語Wikipediaオントロジー」の構築を行っている。

本研究でも軍艦に関する知識の体系化を行うオントロジーの構築において、知識ベースとしてDBpediaを用いる。これにより、既にオーブ

ンデータとして公開されている情報の再利用を行い、多様なデータの活用や、オントロジー構築に要するコストの削減を目指す。また、本研究では軍艦オントロジーの可視化を行い、軍艦に関する多角的な情報提供と、平賀譲 DA が公開する資料への多角的なアクセスの実現を目的とする。

## 2. 軍艦オントロジーの構築

本研究で構築する軍艦オントロジーと平賀譲 DA の目録データの関係を図 1 に示す。以下では、DBpedia を用いた軍艦オントロジーの構築と、平賀譲 DA の目録データとの関連付けに分けて説明する。なお、本研究における「軍艦オントロジー」とは、大日本帝国海軍が所持していた軍艦に関するオントロジーとする。

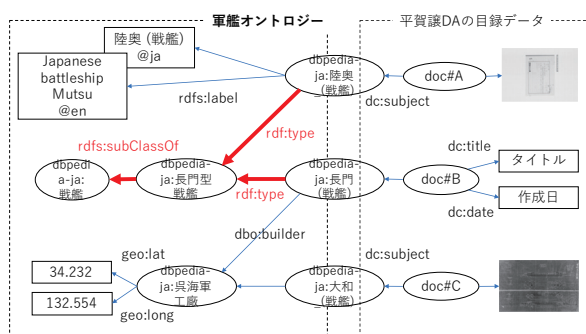


図 1 軍艦オントロジーの概要  
Figure 1 Overview of Warship Ontology.

### 2.1. DBpedia を用いた軍艦オントロジーの構築

まず、構築する軍艦オントロジーの要件定義を行う。本研究では、書籍『日本の軍艦』[7]の「艦艇一覧表」で使用されている 15 項目を、軍艦に関する情報を記述するための要件項目として定めた。具体的には、表 1 の「項目名」列に示す上位 15 項目である。艦名や艦種の他、排水量や全長、速力等の要目、および起工年月日や建造所等の情報を記述するための項目から構成される。

次に、DBpedia を用いたオントロジーの構築にあたり、DBpedia オントロジーとして提供されているプロパティと先の要件項目とのマッピングを行った。この結果、表中の「プロパティ」列に示すように、15 項目のうち、「艦種」を除く 14 項目が DBpedia オントロジーで提供されていることがわかった。さらに、Wikipedia の記事に関する特徴を考慮し、「概要文」と「サムネイル」についても、構築するオントロジーの項目として追加した。

表 1 軍艦オントロジーに関する項目一覧  
Table 1 Properties to describe Warship Ontology.

項目名	プロパティ	値の例
1 艦名	rdfs:label	Japanese battleship Mutsu (en)
2 艦種		
3 排水量 (t)	dbo:shipDisplacement	3272000000.000000 (xsd:double)
4 全長 (m)	dbo:length	215.800000 (xsd:double)
5 最大幅 (m)	dbo:shipBeam	29.020000 (xsd:double)
6 吃水 (m)	dbo:shipDraft	9.080000 (xsd:double)
7 推進	dbp:shipPower dbp:shipPropulsion	*4 shafts *4 × steam turbines
8 速力 (knot)	dbo:topSpeed	49.078000 (xsd:double)
9 兵装	dbp:shipArmament	*4 × twin 41 cm guns
10 装甲	dbp:shipArmor	*Deck: + + *Turrets: *Barbettes:
11 搭載機	dbp:shipAircraft dbp:shipAircraftFacilities	3 (xsd:integer)
12 起工年月日	dbo:layingDown	1917-08-28 (xsd:date)
13 進水年月日	dbo:shipLaunch	1919-11-09 (xsd:date)
14 竣工年月日	dbo:commissioningDate	1920-11-25 (xsd:date)
15 建造所	dbo:builder	dbr:Kure_Naval_Arsenal
* 概要文	rdfs:comment	Nagato was a super-dreadnought battleship built for the Imperial Japanese...
* サムネイル	dbo:thumbnail	wiki-commons:Special:FilePath/...

一方、DBpedia オントロジーにおいて、艦種を記述するためのプロパティが提供されていないことがわかった。したがって、文献[7]を参考として、軍艦と艦種の間を人手で構築した。この人手で付与したリンクは、図 1 中の太線で示すリンクである。関連付ける艦種の URI については、DBpedia が提供する URI を用いた。これは「戦艦」や「巡洋艦」などの艦種についても、Wikipedia の記事として存在するためである。このため、DBpedia において、艦種はインスタンスとして扱われるが、本オントロジーではクラスとして定義した。したがって、艦種間の関係 (ex. 「戦艦」と「長門型戦艦」) は rdfs:subClassOf、艦種と軍艦の関係 (ex. 「長門型戦艦」と「長門(戦艦)」) については rdf:type をそれぞれ用いて関連付けを行なった。このような軍艦階層の記述を行うことにより、単純な軍艦の階層関係の表現に加え、階層関係に基づく推論処理の適用が可能となる。

これらのプロセスを通じ、123 件の軍艦のリソースから構成されるオントロジーを構築した。知識ベースとして DBpedia を用いることにより、既に公開されている軍艦に関する様々な情報を再利用することができる。例えば、Wikipedia の記

事は、複数の言語で記述されることが多い。そのため、例えば戦艦「陸奥」等の軍艦のリソースについて、DBpedia が提供する日英二言語による軍艦名 (rdfs:label) や概要文 (rdfs:comment) を利用することができる。また、図中の例では、ある軍艦に関する建造所がリソースとして与えられ、さらに建造所の緯度経度の情報 (geo:lat, geo:long) が与えられている。この情報を利用することにより、建造所別の軍艦に関する情報提供等が可能となる。

## 2.2. 平賀譲 DA の目録データとの関連付け

次に構築したオントロジーと平賀譲 DA の目録データとの関連付けを行う。目録データは、5463 件の資料に関する書誌情報であり、具体的には表 2 の「項目名」列に示す項目から構成される。

前節で構築したオントロジーとの関連付けを行うため、まず目録データの RDF データへの変換を行う。表中の「プロパティ」列に示す値は、各項目をマッピングした結果を示す。基本的に Dublin Core や DC-NDL 等の既存語彙を使用し、「カード目録」や「文書記述形式」等の平賀譲 DA で独自に使用される項目については、独自語彙 hp (hiraga property の意) を定めてマッピングを行った。

表 2 平賀譲デジタルアーカイブの目録データ  
Table 2 Catalog data of Yuzuru Hiraga Digital Archive.

項目名	値の例	プロパティ
1 表題	"KONGO,FUSO,ISE AND NAGATO". CALCULATION OF STRESS ON GIRDER TO SUSPEND THE RUDDER.	dc:title
2 作成年月日	1918-02-18	dc:date
3 枚数	8	dc:format
4 資料ID	10170201	dc:identifier
5 所蔵機関	東京大学柏図書館	dcndl:holdingAgent
6 登録年月日	2007	dcndl:dateDigitized
7 サムネイル	http://gazo.dl.itc.u- tokyo.ac.jp/hiraga...jpg	foaf:thumbnail
8 カード目録	(舵資料)等 設計図、計算書、メモ書き、印刷物など	hp:card
9 カテゴリ	軍艦構造:舵	hp:category
10 文書記述形式	英文タイプ・図面青焼き・罫紙ペン	hp:format
11 文書タイプ	報告書・メモ	hp:type
12 キーワード	金剛, 扶桑, 伊勢, 長門	dc:subject

次に、RDF によって表現された目録データと軍艦オントロジーの関連付けを行う。平賀譲 DA の目録データにおいては、表中の最下部の行に示すように、各資料に関する軍艦名がキーワードとして与えられている。本研究では、この文字列として与えられている軍艦名と、オントロジーにおける軍艦のリソースとの関連付けを行う。このようなエンティティリンキング、または Wikification のタスクを自動化する手法[8]も存在するが、本研究では単純な文字列のマッチングによる関連付けを行う。この理由として、関連付けの対象となる軍艦のリソースが 123 件に限定

されており、また軍艦のラベルに重複がないためである。具体的には、目録データにおいて文字列として与えられている軍艦名 (ex. 「長門」) と、オントロジーにおける軍艦リソースの日本語ラベル (ex. 「長門 (戦艦)」) の編集距離を算出し、最も類似度が高いラベルを持つ軍艦のリソースの URI (ex. dbr:Japanese\_battleship\_Nagato) を dc:subject の値として与えた。

この結果、図 1 に示すように、軍艦のリソースを介して、オントロジーと目録データが関連付けられ、軍艦オントロジーが提供する関係性を用いた資料へのアクセスが可能となる。

## 3. ケーススタディ

構築したオントロジーの具体的な活用事例について述べる。本ケーススタディは、軍艦オントロジーの可視化による軍艦に関する多角的な情報提供と、平賀譲 DA が公開する資料への多角的なアクセスを可能とすることを目的とする。

図 2 に本ケーススタディで使用するアプリケーションの関係を示す。図左下部は、目録データとともに画像データを公開する平賀譲 DA である。図上部は Virtuoso を用いて構築した RDF ストアを示す。本 RDF ストアにおいて、RDF データに変換された平賀譲 DA の目録データと、DBpedia を用いて構築した軍艦オントロジーが格納される。図右下部は、SPARQL Endpoint を介して RDF ストアにアクセスするアプリケーションであり、抽出したデータの可視化等を行う。本アプリケーションから平賀譲 DA が公開する資料へリンクすることにより、目録データを用いた検索とは異なる観点による資料へのアクセスを実現する。

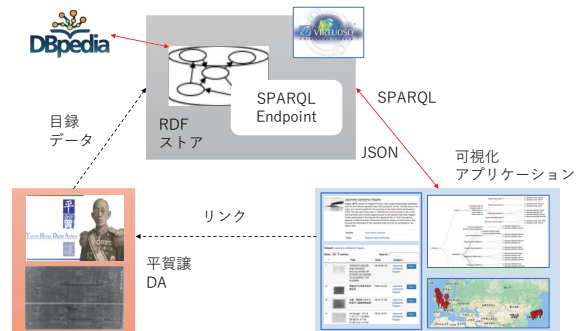


図 2 アプリケーションの関係図  
Figure 2 Relationship of applications.

### 3.1. 軍艦オントロジーの可視化による多角的な情報提供

前章で構築したオントロジーを用いて、軍艦に関する多角的な情報提供を行った事例について述べる。具体的には、軍艦の建造所に関する位置情報、起工年月日や進水年月日等の建造年月日に関する情報、および軍艦と艦種に関する階層

関係等を用いた情報の可視化を行った。この軍艦オントロジーの可視化インタフェースの関係を図 3 に示す。



図 3 軍艦オントロジーの可視化例  
Figure 3 Visualization of Warship Ontology.

図左上部に示すインタフェースは、軍艦名や艦種、起工年月日や進水年月日等の建造年月日に関する情報とともに、軍艦の一覧を表示するインタフェースである。本オントロジーが提供する軍艦の索引を提供する役割を持つ。

図左下部は軍艦の起工年月日を始点とし、タイムライン上にマッピングした軍艦史年表を示す。戦艦、巡洋艦、航空母艦などの艦種別に色を変えて表示する。また、艦種毎に表示・非表示を切り替えるための機能を提供する。この機能を用いることにより、一例として「海軍休日」という軍艦史における一つの事象を本インタフェース上で再現することができた。海軍休日とは、第一次世界大戦終了後のワシントン海軍軍縮条約の締結(1922年)からロンドン海軍軍縮条約の失効(1936年)までの、軍艦の建造が国際協定によって制限された期間のことであり、戦艦(主力艦)の建造が禁止された。この事象について、タイムライン上で艦種として戦艦のみを表示することにより、当該期間において戦艦が建造されていないことを視覚的に確認することができた。なお、本インタフェースの構築には、Javascript ライブラリである vis.js を利用した。

図右上部は艦種と軍艦の階層関係をツリー形式で表現した軍艦階層図である。この軍艦階層の活用については、次節の平賀譲 DA における資料検索への活用事例において詳述する。

図右下部は建造所の緯度経度の情報を用いて、建造所を地図上にマッピングしたインタフェースである。各建造所には、そこで建造された軍艦の数を表示するマーカーが表示される。マーカーを選択することにより、建造所の概要と建造された軍艦の一覧が表示される。また、本インタフェースにおいて、軍艦の建造年月日に関する情報を用いることにより、軍艦の建造所の変遷を可視化することができる。例えば、図 4 は 1904 年に起きた日露戦争前後の国内外における建造件数を比較したものである。日露戦争以前では英国を中心とした軍艦の建造件数が多いのに対し、それ以降は国内における建造が増加していることを視覚的に確認することができる。これは、日露戦争の直後から日本の軍艦の国産化が始まったという軍艦史を表現することができたことを示す。

さらに、これらの各種インタフェースから、図 3 中央部に示すインタフェースに遷移し、各軍艦に関する情報を閲覧することができる。また、当該戦艦に関する平賀譲 DA の資料も合わせて表示する。

以上のように、軍艦オントロジーを地図や年表、階層図等と組み合わせて可視化することにより、軍艦に関する情報を多角的な観点で提供することが可能となった。

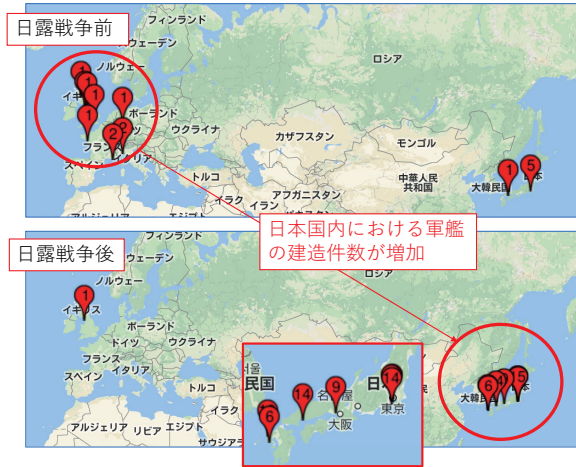


図 4 日露戦争前後の国内外における建造件数の変化  
Figure 4 Domestic production of Japanese warships after the Russo - Japanese War

### 3.2. 平賀譲 DA における資料検索への活用

次に、軍艦オントロジーの平賀譲 DA における資料検索への活用例について述べる。図 5 は軍艦階層に基づく推論を用いた資料検索インタフェースである。



図 5 軍艦階層を用いた推論検索例  
Figure 5 Inference Search with Warship Hierarchy.

本インタフェースは、画面左部に軍艦の階層図が表示される。この階層図において、ある軍艦を選択することにより、選択した軍艦もしくは艦種のインスタンスである軍艦の一覧が右画面上部に表示される。さらに、それらの軍艦をキーワードとして持つ平賀譲 DA の資料一覧が合わせて表示される。図中の例では、軍艦階層図において「長門型戦艦」が選択され、その下位軍艦である「長門」「陸奥」両戦艦に関する資料の一覧が表示されている例を示す。特に、「長門型戦艦」に関しては、後継艦である「加賀型戦艦」の建造計画にかけて、煙突本数が減少するなどの重要な設計変更が見られた。したがって、本例のような同型艦の軍艦に関する資料を一括で検索できることは、有用な検索手段の一つとして考えられる。

その他、前節で述べた建造所に関する情報を活用することにより、例えば「英国（の建造所）で建造された軍艦に関する資料の検索」といった多様なクエリを発行することが可能となる。このような検索は、軍艦名の文字列をキーワードとして提供する従来の平賀譲 DA では実行することができない検索方式である。これらの観点から、構築した軍艦オントロジーが、平賀譲 DA における多角的な資料へのアクセスの実現に寄与したことを確認した。

### 3.3. ケーススタディのまとめ

本ケーススタディで述べた各インタフェースは、SPARQL Endpoint へ発行するクエリの抽出条件の言語を切り替えることにより、日英二言語で表示することができる。これにより、日本語に馴染みのない利用者に対する情報提供、および資料へのアクセスを支援することが可能となる。これらの観点から、本研究における軍艦オントロジーの構築と利用を通じ、平賀譲 DA の課題として取り上げた「多言語対応」および「知識体系の整理」の両者の解決に寄与する手法を提案することができたと考える。

## 4. 考察

### 4.1. DBpedia を用いたオントロジー構築の利点

本研究では、Linked Open Data の一つである DBpedia を用いたオントロジーの構築を行った。既存の知識ベースが提供するデータを再利用することにより、低コストでのオントロジー構築を行った。具体的には、人手による作業は 123 件の軍艦のリソースに対応する艦種の情報追記のみであり、その他の軍艦の建造所や建造年月日、複数言語による軍艦名のラベル等のデータは機械的に抽出・利用した。また、構築したオントロジーを Linked Data として利用することにより、データの可視化等を行うアプリケーションの構築についても、SPARQL Endpoint を用いて簡易に実現することができた。これらの点が、DBpedia を用いたオントロジー構築における利点として考えられる。

### 4.2. 情報の信頼性と網羅性

一方、DBpedia を用いたオントロジー構築の課題として、DBpedia (Wikipedia) が提供する情報の信頼性に加え、情報の網羅性が挙げられる。本研究で取り扱った軍艦のリソースにおいても、建造年月日や建造所に関する情報が欠落しているものが含まれる。構築したオントロジーの正確性および網羅性の検証に向け、人手による情報の追記や修正が必要となる。

この時、特に網羅性の向上を目的として、欠落した情報の追加作業を行う場合、前章のケーススタディで述べたオントロジーの可視化結果の活用を検討する。具体的には、図 3 左上部に示したように、各軍艦のリソースが持つ情報を一覧で表示することにより、情報が欠落している箇所を明らかにすることができる。表 3 は建造年月日や建造所別の網羅率を示す。この表から、例えば、竣工年月日のように網羅率が低い項目について、単なる情報の欠落なのか、もしくは軍縮等により建造が中止となった軍艦に関する情報なのか等、検証すべき項目の見える化が可能となる。

表 3 軍艦オントロジーにおける項目別の網羅率  
Table 3 Coverage by properties of Warship Ontology.

項目	割合	項目	割合
起工年月日	123/123	建造所	103/123
進水年月日	123/123	竣工年月日	35/123
サムネイル	120/123		

#### 4.3. 軍艦オントロジーの拡充

また、上述した情報の信頼性と網羅性の向上に加え、項目の追加等によるオントロジーの拡充を検討している。具体的には、各軍艦が参戦した海戦などの戦歴に関する情報や、軍艦の改修履歴に関する情報の拡充を検討している。

戦歴に関しては、Wikipedia のカテゴリとして、「日露戦争で沈没した軍艦 (dbc: Shipwrecks\_of\_the\_Russo-Japanese\_War)」や「第二次世界大戦で使用された戦艦 (dbc: World\_War\_II\_battleships\_of\_Japan)」等が存在するため、この情報を用いたオントロジーの拡充が可能だと考える。

一方、改修履歴は、Wikipedia において扱いにくい情報である。Wikipedia では基本的に軍艦毎に記事が作成され、改修履歴等はその記事内の Infobox 等に記載される。したがって、DBpedia においては、改修前後で排水量や速力等が変化した軍艦であっても一つのリソースとして扱われ、排水量や速力等の項目に対して改修前後の 2 つの値が与えられる。このようなデータの持ち方は、改修前後の要目比較等に適していない。この課題に対して、改修前後の軍艦を異なるリソースとして表現すること等が必要となる。このような Wikipedia の記事の単位とは異なる粒度でのデータの取り扱いに向け、今後は構造化データの管理を前提とする Wikidata[9]を用いたリソースの追加や情報の補完等を行うことを検討する。

## 5. 結論

本研究では、DBpedia を用いてオントロジーを構築することで、既存の知識ベースが提供するデータを再利用することが可能となり、123 件の

軍艦から構成されるオントロジーを低コストで構築することができた。また、構築した軍艦オントロジーを用いて、軍艦の建造所や建造史、軍艦と艦種に関する階層構造等を日英二言語で可視化するアプリケーションを構築した。本アプリケーションを用いることで、軍艦に関する多角的な情報提供と、平賀譲 DA が公開する資料に対する多角的なアクセスを支援することができた。

今後は本研究で構築した軍艦オントロジーの拡充を行い、研究や展示利用におけるさらなる活用を検討する。

## 参考文献

- [1] 平賀譲デジタルアーカイブ, 入手先 <<http://gazo.dl.itc.u-tokyo.ac.jp/hiraga2/>> (参照 2017-11-15).
- [2] 古崎晃司, 來村徳信, 溝口理一郎: 生物規範工学オントロジーと Linked Data に基づくキーワード探索, 人工知能学会論文誌, Vol. 31, No. 1, pp.1-12 (2016).
- [3] 研谷紀夫: 皇族・華族を対象とする人名典拠情報の構築と Digital Cultural Heritage への活用, アートドキュメンテーション研究, No. 19, pp.36-53 (2012).
- [4] 安田つくし, 三原鉄也, 永森光晴, 杉本重雄: Linked Open Data を用いた同人創作物探索支援のためのメタデータの構築, じんもんこん 2014 論文集, Vol. 2014, No. 3, pp.177-184 (2014).
- [5] 武田英明, 加藤文彦, 大向一輝: Linked Data による分野連携型データベースの枠組み, 情報知識学会誌, Vol. 25, No. 4, pp.283-290 (2015).
- [6] 玉川奨, 香川宏介, 森田武史, 山口高平: 大規模 Linked Open Data のための日本語語彙の構築, 人工知能学会論文誌, Vol. 29, No. 4, pp.386-395 (2014).
- [7] 福井静夫: 日本の軍艦—わが造艦技術の発達と艦艇の変遷, 出版協同社 (1993).
- [8] Davaajav Jargalsaikhan, 岡崎直観, 松田耕史, 乾健太郎: 日本語 Wikification コーパスの構築に向けて, 言語処理学会第 22 回年次大会 (2016).
- [9] 加藤文彦: DBpedia の現在: リンクトデータ・プロジェクト, 情報管理, Vol. 60, No. 5, pp.307-315 (2017).