

日本語 Wikipedia からの セマンティック Web コンテンツの半自動構築

Building Semantic Web Contents Semi-Automatically from Japanese Wikipedia

森田 武史

近年、セマンティック Web コンテンツ構築のための新たな情報資源として Wikipedia に注目が集まっている。本稿では、日本語 Wikipedia における様々なリソース（カテゴリツリー、一覧記事、リダイレクトリンク、Infobox、Infobox テンプレート）から、概念および概念間の関係（Is-a 関係、クラス-インスタンス関係、プロパティの定義域および値域、同義語、インスタンス間関係）を抽出することにより、高精度かつ大規模なセマンティック Web コンテンツを構築する手法を紹介する。

1. はじめに

現在の Web コンテンツは、人間が理解することを目的として構築されているため、ソフトウェアによる理解は困難である。一方、次世代 Web の候補の一つであるセマンティック Web (Berners-Lee, 2001) は、ソフトウェアが意味理解可能な辞書 (オントロジー) に基づいて、Web コンテンツにソフトウェア可読なメタデータを付与することによって、ソフトウェアが Web コンテンツを理解し、推論することを可能にしようという試みである。セマンティック Web の実現により、知的な情報検索やアプリケーションを横断したデータ統合および再利用などが可能となる。しかしながら、ソフトウェアが Web コンテンツを理解するためのセマンティック Web 標準技術である、RDF (Resource Description Framework)¹、RDFS (RDF Schema)²、OWL (Web Ontology Lan-

guage)³ により記述されるセマンティック Web コンテンツ (オントロジーおよびそのインスタンスネットワーク) は、現状では人手によって構築されており、多大なコストを要している。オントロジーの手動構築には、膨大な時間がかかり、間違いを起しやすく、オントロジーの保守や更新が困難という問題がある。そこで、近年、オントロジー工学のコミュニティは、オントロジー開発コストを削減するために、オントロジー学習 (Ontology Learning) と呼ばれる、(半)自動的にオントロジーを構築する手法、方法論、アルゴリズム、ツールなどの研究開発に取り組んできた。特に、フリーテキストからのオントロジー学習に関しては、機械学習、知識獲得、自然言語処理、情報検索など、様々な専門分野の手法を組み合わせた手法がこれまで数多く提案されている (Buitelaar, 2005)。しかし

¹ <http://www.w3.org/TR/rdf-syntax-grammar/>

² <http://www.w3.org/TR/rdf-schema/>

³ <http://www.w3.org/TR/owl-ref/>

ながら、非構造情報資源であるフリーテキストと構造情報資源であるオントロジーの間のギャップは大きく、高精度で、大規模なオントロジーを構築することは困難であるのが現状である。

一方、近年、Web 上の百科事典である Wikipedia が、新たな情報資源として注目を集めている(中山, 2009)。Wikipedia は語彙網羅性、即時更新性に優れており、半構造情報資源であることからフリーテキストと比べてオントロジーとのギャップが小さい。そのため、Wikipedia からのオントロジー学習研究が近年、盛んに行われている。しかしながら、Wikipedia はユーザ参加型という性質上、厳密な体系化が行われていないため、Wikipedia からのオントロジー学習にも、多くの課題が存在している。

本稿では、日本語 Wikipedia における様々なリソースから、概念および概念間の関係を抽出することにより、高精度かつ大規模なセマンティック Web コンテンツを半自動的に構築する手法を紹介する。ここで構築するセマンティック Web コンテンツを本研究では日本語 Wikipedia オントロジーと呼ぶ。また本研究では 2009 年 10 月時点の日本語 Wikipedia を用いている。

2. 日本語 Wikipedia オントロジーの構築

本章では、日本語 Wikipedia から、以下に示す 5 種類の関係を抽出することにより日本語 Wikipedia オントロジーを(半)自動的に構築する手法について述べる。以下の()内は、抽出した関係に対応する、OWL, RDFS, RDF, SKOS⁴ で定義した語彙(クラス及びプロパティ)を示す。

- (1) Is-a 関係 (rdfs: subclassOf)
- (2) クラス-インスタンス関係 (rdf: type)

- (3) Infobox トリプル (owl: Object/DatatypeProperty)
- (4) プロパティ定義域 (rdfs: domain)
- (5) 同義語 (skos: altLabel)

2.1. Is-a 関係構築

2.1.1. 予備実験

Wikipedia には、記事の分類を目的とした階層的なカテゴリが存在する。しかし、下位カテゴリと上位カテゴリとの関係は、性質の継承という観点から捉えた Is-a 関係ではない場合が多々見られる。実際にカテゴリ階層がどの程度 Is-a 関係を持っているかを知るために、日本語 Wikipedia のカテゴリ関係のダンプデータから 1,000 個の標本を抽出し、下位カテゴリと上位カテゴリがどのような関係になっているのか予備実験を行った。その結果、Is-a 関係は 50.9%であった。誤りの例としては、「アニメキャラクター」とその「声優」が階層関係になっているものや、「オリンピックメダリスト」と「オリンピック」といった人物と事象が階層関係になっているものなどが多く見られた。また、クラス-インスタンス関係も全体の 10.1%であり、「スーパーマリオ」や「SONY」といった「ゲーム作品」や「企業」クラスのインスタンスも存在した。

以上の予備実験の結果から、Wikipedia カテゴリ階層をそのまま Is-a 関係に利用する事は困難であるため、本研究では Is-a 関係を以下の 2 つの手法により構築する。

- (1) カテゴリ階層に対する文字列照合
- (2) カテゴリ名と Infobox テンプレートの照合

2.1.2. カテゴリ階層に対する文字列照合

あるカテゴリから相対的に下位に存在するサブカテゴリは、増加した記事を細分化するために作成されるという性質から、上位カテゴリの名称を含む複合語で形成される場合が多い。例えば「原子カー原子力発電所」や「ソ

⁴ SKOS: <http://www.w3.org/TR/skos-reference/>

フトウェアフリーソフトウェア」といった階層である。前者は性質の継承という観点からみた Is-a 関係としては不適切な関係であるが、後者は Is-a 関係に相当する関係となっている。本論文では、Is-a 関係を抽出するためのカテゴリ階層の複合語に対する文字列照合として、「後方文字列照合」と「前方文字列照合部除去」を行う。

後方文字列照合とはカテゴリ階層を構成する親カテゴリ名と子カテゴリ名とを比較し、子カテゴリ名が「任意の文字列+親カテゴリ名」となっているものを抽出する手法である。図1の例では「空港」という文字列の後方文字列照合により、「日本の空港」Is-a「空港」という関係を得ることができる。後方文字列照合は、(Ponzetto, 2007)で既に提案されている手法である。

前方文字列照合部除去とは親カテゴリ名と子カテゴリ名とを比較し、親カテゴリ名と子カテゴリ名で「任意の文字列+の」という部分が先頭から一致しているものを抽出し、照合部を除去する手法である。図2の例では「日本の」という前方文字列照合部を除去することにより、「ゴルファー」Is-a「スポーツ選手」という関係を得ることができる。この手法は、文字列の重複に依存しない Is-a 関係を取得できる点が大きな利点である。

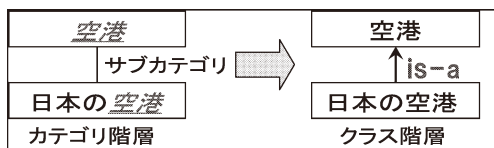


図1：後方文字列照合の例

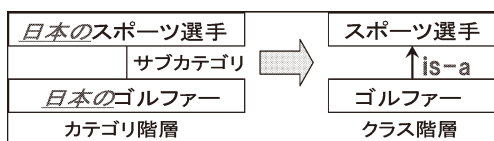


図2：前方文字列照合部除去の例

2.1.3. カテゴリ名と Infobox テンプレートの照合

Infobox は、テーブルを利用して Wikipedia の記事の属性 (Wikipedia では主に「項目」と呼ばれている) と属性値を整理して表示しているもので、記事の中にしばしば掲載されている。ここで使用される項目が、領域ごとにある程度フォーマット化されているということが大きな特徴である。例えば「Java」の記事に掲載されている Infobox には「開発者」や「プラットフォーム」などの項目とそれぞれに対応する値が記述されており、この「開発者」や「プラットフォーム」という項目は、Infobox のテンプレート「プログラミング言語」で定められている。本研究では、各 Infobox の持つ抽象的なテンプレート名と、領域によっては多くの具体的な概念を持つカテゴリ名との関係に着目する。テンプレート名とカテゴリ名の照合を行い Is-a 関係を抽出する。抽出の手順を以下の(1)~(4)に示す。

- (1) カテゴリとテンプレートの情報をデータベースに格納
- (2) カテゴリ名とテンプレート名の単純文字列照合
- (3) 照合したカテゴリ以下に存在するサブカテゴリ名と、照合したテンプレートを持つ記事が所属する全てのカテゴリ名との照合
- (4) 照合によって得たサブカテゴリ名をテンプレート名と Is-a 関係が成り立つとして抽出

図3に、Infobox テンプレートとカテゴリ名の照合による Is-a 関係抽出の具体例を示す。図3は、「楽器」テンプレート、「楽器」テンプレートを用いて作成した Infobox を持つ「ピアノ」と「フルート」記事、「ピアノ」と「フルート」記事が所属するカテゴリ、カ

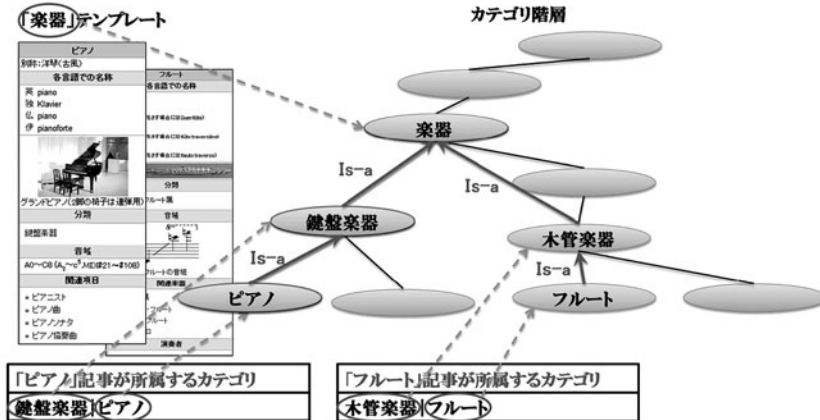


図3：Infobox テンプレートとカテゴリ名の照合

カテゴリツリーとそれらの間の関係を示している。まず、「楽器」テンプレート名とカテゴリツリーの照合を行い、楽器カテゴリを同定する。次に、楽器カテゴリのサブカテゴリ名と「ピアノ」および「フルート」記事が属するカテゴリ名を照合する。その結果、「鍵盤楽器」Is-a「楽器」、「ピアノ」Is-a「鍵盤楽器」、「木管楽器」Is-a「楽器」、「フルート」Is-a「木管楽器」の4つのIs-a関係が抽出できる。ここで、「ピアノ」Is-a「鍵盤楽器」と「フルート」Is-a「木管楽器」は、文字列照合では抽出できないIs-a関係である。

以上の手順を行うことによって、2.1.2節で述べた文字列の特性を利用した「カテゴリ階層に対する文字列照合」では抽出することのできなかったIs-a関係を抽出できる。

それに伴い、正しくないIs-a関係を多く持つWikipediaカテゴリツリーの洗練が可能になると考えられる。

2.2. クラス-インスタンス関係構築

一覧記事とは、ある基準に従って、関連する物事が列挙された記事である。例えば、「言語の一览」記事には、世界の言語が列挙されている。一覧記事は記述形式が統一されているため、記事名をクラス、記事内に列挙された物事をインスタンスとみなして、スクレイ

ピングによりクラス-インスタンス関係を抽出することができる。本研究では、以下に示すスクレイピングの手順に基づきWikipediaの一览記事のソースからクラス-インスタンス関係を抽出する。

- (1) 目次、見出しの除去
- (2) 一覧記事の説明に使用される“*”行を除去
- (3) “*～一覧”と同じ目次見出しの下位にある“*”行を除去
- (4) 不要な目次見出し下位の“*”行を除去
- (5) 不要な“*”行を除去
- (6) 不要な年号記述部分を除去
- (7) “*”行からのインスタンスの抽出

上記の(1)から(6)までの処理を行った後、“*”行の中でどの部分がインスタンスを表す文字列であるかを特定する六つのパターンを作成した。図4が“*”行中でインスタンス箇所を特定する6つのパターンである。

以上の手順に従って不要な部分をスクレイピングにより排除し、最終的に、記事名と抽出した項目名の関係をクラス-インスタンス関係として抽出する。

1 2.start↓	1 2.start↓	15 167.start↓
2 社会学者↓	2 社会学者↓	16 物理単位↓
3 76↓	3 76↓	17 6941↓
4 ↓	4 ↓	18 ↓
5 ==[[日本]]==↓	5 ==[[日本]]==↓	25 ===長さ===↓
6 ↓	6 ↓	26 * [[メートル]]: m↓
7 ===あ===↓	7 ===あ===↓	27 * [[オングストローム]]: 1 Å = 10<su
8 * [[青井和夫]]↓	8 * [[赤川学]] (セクシュアリティ研究)	28 * [[ポイント]]: ビックポイント 1 pt = 1/72 i
9 * [[浅野智彦]]↓	9 * [[鯉坂学]] (同郷団体の研究) ↓	
10 * [[阿閉吉男]]↓		

① * [[]]

② * [[]] () () ...

③ * [[]]:

0 2680.start↓	1 === 茶道家 ===↓
1 武庫川女子大学の人物↓	2 * [[千利休]] (安土桃山時代の茶人、侘び茶の完成者) : 堺市↓
2 505150↓	3 * [[武野燭燭]] (室町時代後期の茶人、富商) : 堺市↓
3 ↓	
4 == 教職員 ==↓	
5 * [[多田道太郎]] - 仏文学者↓	
6 * [[橋爪静夫]] - 健康・スポーツ科学科教授	

④ * [[]]-

1 61.start↓
4 賞↓
5 4417↓
6 ↓
7 === 学術 ===↓
8 * [[ウルフ賞]] (ウルフ賞とも) - 多部門 (科学・芸術) ↓
9 * [[アルバート・ラスカー医学研究賞]] (ラスカー賞) - 医学

⑤ * [[]] () () ... :

⑥ * [[]] () () ... -

図4：“*” 行中でインスタンス箇所を特定するパターン

2.3. Infobox トリプル

Infobox を有する「記事-項目-値」という三つ組は、「インスタンス-プロパティ-プロパティの値」という三つ組と捉えることができる。そのため、Wikipedia ダンプデータから直接トリプルとして記事タイトルごとのプロパティを抽出できるが、いくつかの問題が存在する。まず、Infobox には記事の種類ごとにテンプレートが存在し、かつ英語版 Wikipedia のテンプレートを利用できるという問題点がある。これは MediaWiki を用いて日本語版と英語版で完全互換性をとっており、そのため記事の執筆者が簡単に編集できるための措置であるが、ダンプデータからトリプルを抽出する際には英語表記と日本語表記でプロパティが別のもとなってしまう。図6の例では、記事ソース内には「Genre」という単語が述語になっているが、実際の記事では「ジャンル」に変換される。このため、記事ソースから直接 Infobox トリプルを抽出すると、「Genre」プロパティとしてそのまま抽出してしまう。次に、全てのプロパティ

の値をリテラルとしてとってしまうと、プロパティの値がデータ値となるのかインスタンスとなるのかの区別が出来ず、プロパティの種別がわからないという問題がある。図5の例では、ジャンルや開発元プロパティの値は owl: ObjectProperty によりインスタンスと関連付けるべきであるが、人数プロパティの値は owl: DatatypeProperty によりリテラルと関連づけるべきである。

以上2つの問題に対応するため、以下の(1)~(4)の手順でプロパティの抽出を行う。

- (1) 記事ごとに Infobox とそのテンプレートの情報をデータベースに格納
- (2) JavaWikipedia API (Bliki engine)⁵ を用いて HTML ソースに変換することで、Infobox を日本語記述で統一して抽出し、Infobox から「インスタンス-プロパティ-リテラル」の形でデータベースに格納

⁵ <http://code.google.com/p/gwtwiki/>

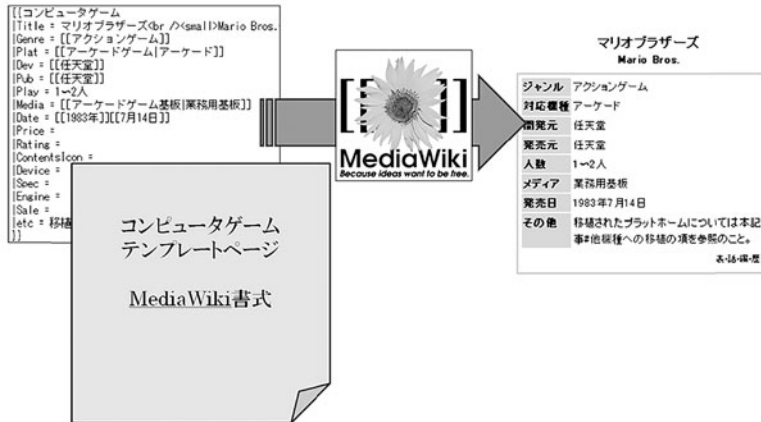


図 5 : MediaWiki を用いた Infobox テンプレートの一例

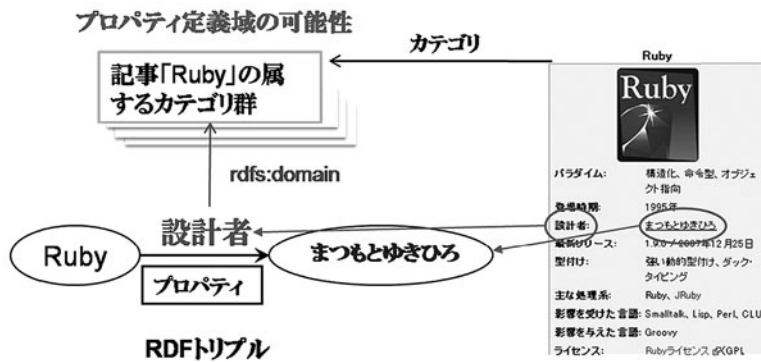


図 6 : Infobox と RDF トリプル, カテゴリと定義域の対応

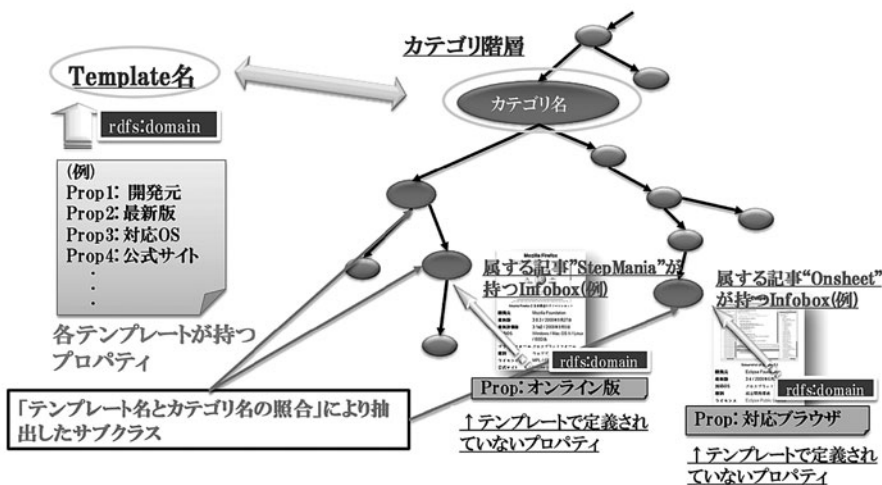


図 7 : テンプレートで定義されていないプロパティ定義域の抽出

- (3) 40種類のInfoboxテンプレートにおけるプロパティについてモデリングを行う。
- (4) (3)とデータベースに格納したトリプルから新たなトリプルを抽出

(3)においてモデリングとは、各テンプレートのプロパティの目的語がインスタンスになるかリテラルになるか、また、目的語がリテラルになる場合にはそのデータ型を記述することを意味する。この際用いるInfoboxテンプレート数を40と指定したのは、2009年10月のWikipediaダンプデータを使用し、事前に抽出した際にInfoboxの総数が約20万2,000個だったのに対し、出現頻度が高かった上位40種類のInfoboxテンプレートで約14万6,000個(約72%)のInfoboxのモデリングが行えたためである。

以上の手順より、7割程度のInfoboxに対して適切にInfoboxトリプルが抽出できると考えられる。

2.4. プロパティ定義域抽出

2.3節で述べたInfoboxトリプルにおける主語は、記事名をインスタンスとして据えていた。そのため、主語である記事が属するカテゴリを調べることで、プロパティの定義域を定義できる可能性がある。図6は、記事「Ruby」が属するカテゴリがInfoboxトリプルにおける「設計者」プロパティの定義域として定義できる可能性のあることを表している。

本論文では以下に述べる手法でプロパティ定義域の抽出を行う。まず、Infoboxテンプレート名を、Infoboxが持つ各プロパティの定義域として抽出する。次に、2.1.3節で述べた「カテゴリ名とInfoboxテンプレートの照合」により得たIs-a関係として正しいサブカテゴリを、テンプレートの持つ各プロパティの定義域として対応付ける。さらに、テンプレ

レートで定義されていないプロパティの定義域抽出を試みる。

実際に記事に記載されたInfoboxに登場するプロパティは、テンプレートで定義されているプロパティ以外のものが使用されるケースが多数存在する。例えば、「有機化合物」というテンプレートで定義されているプロパティは「構造式」、「形状」、「沸点」など合計21あるが、実際の記事に掲載されているInfoboxのソースから収集したプロパティは、「揮発性」、「臭気」、「蒸気圧」などテンプレートで定義されていないものが多く存在し、合計33のプロパティを持つ。

そこで、本提案手法では、2.1.3節で述べた手法により得たサブカテゴリと、そのカテゴリに属する記事が持つInfoboxテンプレートで定義されていないプロパティとの関係に着目する。Infoboxテンプレートで定義されていないプロパティは2.1.3節で述べた手法により得たサブカテゴリのいずれかをドメインとする可能性があり、これらを対応付けることで、各プロパティの定義域として最上位の概念であるテンプレート名が得られるだけではなく、より具体化し、ドメインに特化したプロパティおよび定義域の抽出が可能になると考えられる。図7がプロパティ定義域の抽出手法の全体像である。記事「StepMania」にはテンプレートに定義されていない「オンライン版」というプロパティが存在する。2.1.3節で述べた手法により得たサブカテゴリと、この記事が属するカテゴリを比較することで、新たに限定した定義域が抽出できる。

2.5. 同義語抽出

Wikipediaにはリダイレクトという機能が存在する。リダイレクトとは、あるページを表示した際に既存の別のページへ自動的にリンクさせる機能である。リダイレクト元の記事名とリダイレクト先の記事名との関係(リダイレクトリンク)は「音楽家」と「ミュー

ジション」など同義語の関係にあることが多く、Wikipedia のリダイレクトリンクを利用することで同義語の抽出が可能となる。

3. 評価と考察

3.1. 日本語 Wikipedia オントロジーの全体像

表 1 に日本語 Wikipedia オントロジー (JWO) の Is-a 関係、クラス-インスタンス関係、Infobox トリプル、プロパティ定義域、同義語の各関係数および正解率の 95% 信頼区間を示す。

Is-a 関係数は約 16,000、正解率は 90% 以上で、高精度となっている。

クラス-インスタンス関係は 421,989 もの関係抽出し、インスタンス数も 323,024 と多く、また正解率も 97.2% と高い。しかし、これらは一覧記事から抽出したインスタンスであり、Wikipedia の記事数が約 64 万記事 (2009 年 10 月時点) あることを考えるとさらに多くのインスタンスを抽出できる可能性がある。

プロパティについては、4,530 プロパティについて、6,485 のプロパティ定義域が抽出できた。また、正解率も約 95% と高精度であった。Infobox トリプルについても、約 140 万もの関係が、正解率約 95% と高精度で抽出できた。

3.2. 日本語 Wikipedia オントロジーの特徴

JWO と代表的な既存汎用オントロジーである、日本語語彙体系 (池原, 1997) および

日本語 WordNet (Bond, 2009) との比較評価を行った。表 2 に JWO と既存汎用オントロジーの比較例を示す。表 2 中の「JWO」は本論文で構築した日本語 Wikipedia オントロジーを、「W」は日本語 WordNet を、「N」は日本語語彙体系をそれぞれ表している。また、「パス」はいくつかの主要クラスからリーフまでのパスを表している。

JWO の特徴として、表 2 の例の「ジャズ・ギタリスト」や「イギリスの空対空ミサイル」クラスのように、特定分野に特化して詳細なクラス階層を定義している点あげられる。さらに、JWO のクラスは膨大なインスタンスを持っている。これらは、他の既存汎用オントロジーにはない、JWO の特徴である。

表 2 の「事物-人物」の関係のように、構築した JWO は他と比べて上位概念が不足していることが分かる。これは、Wikipedia のカテゴリが 9 種の主要カテゴリ (「学問」、「技術」、「自然」、「社会」、「地理」、「人間」、「文化」、「歴史」、「総記」) から分類されているために構築したオントロジーも主要カテゴリをルートとした階層関係になっているためである。また、「人物」から「ギタリスト」クラスまでのパスの深さが浅いことから、中間概念が不足していることもわかる。

4. おわりに

本稿では、日本語 Wikipedia をリソースとして、オントロジー学習による大規模かつ汎用的な日本語オントロジーの構築手法を紹介した。本稿では、すべて紹介することができ

表 1：日本語 Wikipedia オントロジーの関係数と正解率

関係の種類		関係数	正解率
Is-a 関係	文字列照合	12,558	93.1±1.51%
	Infobox テンプレート照合	3,782	95.6±1.09%
クラス-インスタンス関係		421,989	97.2±1.02%
Infobox から得られるトリプル		1,485,751	95.8±1.79%
プロパティ定義域		6,485	95.4±1.22%
同義語		106,671	67.0±2.90%

表2：オントロジー比較の例

クラス	オントロジー	パス
人(音楽家)	JWO	事物-人物-音楽家-演奏家 -ギタリスト-ジャズ・ギタリスト
	W	もの-全般-生き物-生物-人-エンターテイナー -公演者-ミュージシャン-ギタリスト
	N	名詞-具体-主体-人-人職業地位役割 -人職業-人専門的技術的職業-芸術家-音楽家
無生物(兵器)	JWO	文化と歴史-出来事-政治-行政-軍事-兵器-航空兵器 -空対空ミサイル-イギリスの空対空ミサイル
	W	もの-全般-出土品-機器-機器-装甲-ミサイル-空対空ミサイル
	N	"兵器" の類はなし
抽象物(過去)	JWO	Wiki "過去" は未定義
	W	属性-時-古-過去
	N	名詞-抽象-抽象的關係-時間-非暦日-現在過去未来-過去

なかったが、関連研究として、本稿で紹介した手法のより詳細な評価についてまとめた論文(玉川, 2010), セマンティクス(上位・下位, 対称, 推移など)を考慮したプロパティ抽出に関する研究(玉川, 2011), プロパティ継承を考慮したプロパティ付きクラス階層構築の研究(Morita, 2012), 日本語 Wikipedia オントロジーの上位階層の不足を補うために日本語 Wikipedia オントロジーと日本語 WordNet を統合する研究(森田, 2012)などを発表している。また、最新版の日本語 Wikipedia オントロジーは Web 上で公開している⁶。

参考文献

- [1] T. Berners-Lee, J. Hendler, and O. Lassila (2001), "The SemanticWeb," Scientific American, pp.34-43
- [2] Paul Buitelaar, Philipp Cimiano, Bernardo Magnini (Eds.) (2005), Ontology Learning from Text: Methods, Evaluation and Applications Frontiers in Artificial Intelligence and Applications Series, Vol.123, IOS Press

- [3] 中山浩太郎, 伊藤雅弘, Erdmann, M., 白川真澄, 道下智之, 原 隆浩, 西尾章治郎(2009): Wikipedia マイニング近未来チャレンジキックオフ編, 人工知能学会論文誌, Vol.24, No.6, pp.549-557
- [4] Simone Paolo Ponzetto, Michael Strube (2007): Deriving a Large Scale Taxonomy from Wikipedia, Proc. AAAI-2007, AAAI Press, pp.1440-1447
- [5] 池原 悟, 宮崎正弘, 白井 諭, 横尾昭男, 中岩浩己, 小倉健太郎, 大山芳史, 林 良彦 (1997): 日本語語彙体系, 岩波書店
- [6] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi and Kyoko Kanzaki (2009): Enhancing the Japanese WordNet, in The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP
- [7] 玉川 奨, 桜井慎弥, 手島拓也, 森田武史, 和泉憲明, 山口高平(2010), "日本語 Wikipedia からの大規模オントロジー学習", 人工知能学会論文誌 論文特集「2009年度全国大会近未来チャレンジ」Vol.25 No.5 pp.623-636
- [8] 玉川 奨, 森田武史, 山口高平(2011), "日本語 Wikipedia からプロパティを備えたオン

⁶ <http://www.wikipediaontology.org/>

- トロジーの構築”, 人工知能学会論文誌 特集論文「近未来チャレンジ」Vol.26 No.4 pp. 504-517
- [9] Takeshi Morita, Yuka Sekimoto, Susumu Tamagawa, Takahira Yamaguchi (2012), “Building up a Class Hierarchy with Properties from Japanese Wikipedia”, 2012 IEEE/WIC/ACM International Conference on Web Intelligence
- [10] 森田武史, 玉川 奨, 山口高平 (2012), “オントロジーアライメントを用いた日本語 Wikipedia オントロジーと日本語 WordNet の統合”, 人工知能学会学会, 第 28 回セマンティックウェブとオントロジー研究会