

# Wikipedia カテゴリオントロジー

## - Wikipedia カテゴリの知識工学的観点からの活用を目指して -

### Wikipedia Category Ontology

#### - Utilization of Wikipedia category from the view point of knowledge engineers

吉岡真治<sup>1\*</sup>  
Masaharu Yoshioka<sup>1</sup>

<sup>1</sup> 北海道大学大学院情報科学研究院

<sup>1</sup> Faculty of Information Science and Technology, Hokkaido University

**Abstract:** Wikipedia is one of the largest knowledge resources available via the Internet. Several methods have been proposed for extracting knowledge/information from Wikipedia, but methods based on extracting knowledge from the Wikipedia category structure can not utilize whole part of its structure because of the complexity of the relationships between the various categories. In this paper, we briefly review our previous researches on analyzing Wikipedia category and introduce a framework called “Wikipedia Category Ontology” (WCO) that aims to act as a basis for interpreting the Wikipedia category structure. It is based on a classification of category types and relationship types and available online in the form of Linked Open Data at <http://wcontology.org/>. We also demonstrate the system by using Wikipedia category analysis scenarios.

## 1 はじめに

現在、深層学習を中心とした機械学習に基づく自然言語処理技術の進展により、様々な単語の意味などを、大量の文書データから学習することが可能になってきている。しかし、このような形で作成される意味表現は、学習するデータに依存するため、データの偏りなどの影響を受ける。これに対し、概念階層に基づく分類のように、人間が言葉の意味役割を記述的な知識として付与するような技術は、訓練データに依拠しないだけでなく、説明可能な人工知能を作るための基盤となることが期待される。このような観点から、大規模な記述的な知識を Knowledge Graph として構築し活用する研究が盛んになってきている。

このような Knowledge Graph の多くは、Linked Open Data のデータを活用して作られており、中でも多くのボランティア編集者によって編纂されている Web 上の百科事典である Wikipedia<sup>1</sup> は、重要な情報源である。

この Wikipedia は、通常の紙媒体の百科事典と異なり、Template を用いた半構造化された記事の記述や、カテゴリによる分類、リンク構造に基づく記事の関係の明記などを行うことが可能という特徴を持っている。特に、半構造化された記事の形式を活用して、記事の対象となる固有物に関するメタデータの抽出などを行う DBpedia[1] を中心としたデータベースは、Linked Open Data (LOD) の中心的な役割を果たしている。

一方、Wikipedia カテゴリによる分類を活用した研究としては、カテゴリ階層の情報から、概念の上位下位関係の情報を抽出し利用する日本語 Wikipedia オントロジー [2]、カテゴリの情報を WordNet のような外部の概念階層の知識を用いて整理してオントロジーを構築する YAGO[3] の研究などがあるが、カテゴリ階層のうち、特定のパターンを有するものだけを選択的に利用するといった部分的な情報の利用に留まっている。

また、Wikipedia カテゴリ階層構造は、意味的な類似性を図るための情報源としても活用されてきた [4, 5, 6, 7, 8]。しかし、これらの研究では、ベースラインの手法と比較した有用性は示されているものの、複雑な

\*連絡先：北海道大学大学院情報科学研究院  
〒064-0806 札幌市北区北 14 条西 9 丁目  
E-mail: yoshioka@ist.hokudai.ac.jp

<sup>1</sup><http://en.wikipedia.org/>

Wikipedia の階層構造を適切に利用できているわけではないため、これらの手法が発展的に利用される状況にはなっていない。

また、近年の Wikipedia を Knowledge Graph として活用しようとする研究 [9, 10, 11, 8] においては、Wikipedia カテゴリの複雑さを簡単にするために、特定のパターンを満す関係だけを抽出するようなアドホックな方法が用いられている。

このような問題が発生する原因を、知識工学的な観点から考察すると、Wikipedia カテゴリの階層関係には推移律が成り立たない関係が多く含まれること、場合によっては、ループ構造が存在するといった様々な問題点があげられる。

この問題に対し、我々は、Wikipedia カテゴリに関する編集指針を踏まえた分析を行うと共に、そのカテゴリ階層を知識工学的な観点から再構成するための基盤となる Wikipedia カテゴリオントロジー (WCO) を提案してきた。

本論文では、我々が公開している LOD のリソースである WCO の背景となるこれまでの研究について紹介すると共に、WCO の具体的な利用方法を示す。

## 2 Wikipedia カテゴリの特徴

### 2.1 Wikipedia カテゴリの編集方針

Wikipedia カテゴリの知識工学的な観点からの活用の議論をするためには、Wikipedia カテゴリがどのような編集方針で作成されているかを正確に理解することが必須である。

そのため、まず、これまでの Wikipedia カテゴリの分析の研究 [12, 13] でも参照している Wikipedia におけるカテゴリの編集方針が記されているページある「Wikipedia:カテゴリの方針」<sup>2</sup>における、「カテゴリの定義」(図 1)と「カテゴリの構造」(図 2)を示す。

ここでポイントとなるのは、図 2 に関する包含関係に関する記述が親子関係に限定されている点である。そのために、推移的に先祖のカテゴリにたどったときの包含関係について、その包含関係が成立することについて、議論がされていないことが分かる。

また、Wikipedia のカテゴリの構造を複雑にしている原因に、分割として機能するカテゴリ (図 3) がある。これは、Wikipedia を閲覧する人の利便性に考慮した基準であり、サッカークラブやアルバムといった多数の名付き実体を有する分類に対して、適切なサイズのカテゴリに分割するために様々な基準を導入することで、分割されたカテゴリである。多くの場合、このカテゴリは、分割の対象となったカテゴリの名前と、分

1. カテゴリは第一義として、「分類」を示すものです。  
「xx は YY のひとつである」と言うことができれば、「分類」を示すと言えます。項目 xx はカテゴリ YY に属するべきです。反例として、北朝鮮と韓国は関連がありますが、どちらかがどちらかを包含する関係ではありません。
2. 上記に加えて、ウィキペディアのカテゴリとしては「関連が深いキーワード」を示すことができます。  
「分類」より「キーワード」を指向しているカテゴリも存在します。記事 xx が「YY 関連用語」であるという意味合いでカテゴリ YY に属することが期待される場合があります。例として、学術用語と Category:学問の関係など。この場合も、カテゴリはより上位の概念であることが求められるため、逆の関係ではあり得ません。
3. また、カテゴリはウィキペディアの骨組みの意味を持ちます。  
カテゴリ機能の普及によって、カテゴリの構造がウィキペディアの全体構造を示すこととなりました。カテゴリ同士の関係もウィキペディア全体を意識した一貫性や無矛盾性が求められ、よいカテゴリ構造を作ることが、わかりやすいウィキペディアを作ることにつながります。似た意味合いのカテゴリや大きく重複するカテゴリがある場合は、なるべく内容をすり合わせ、統合を検討しましょう。併存させる場合も、明確な使い分けの方針を決めましょう。そうしなければ混乱が永続することになります (例:「文房具」と「事務用品」など)。

図 1: Wikipedia:カテゴリの方針#カテゴリの定義

多くのカテゴリは一つ以上の親カテゴリを持ちます。例えば、Category:日本の作家は Category:各国の作家と Category:日本の人物 (職業別) の両方に含まれています。あるカテゴリを他のカテゴリのサブカテゴリとする場合、前者のカテゴリの内容が (ある程度の例外はありますが) 後者のカテゴリの内容として含まれるものであることを確認してください。カテゴリの上下関係は親子関係であり、ループ構造にならないように注意してください。ある二つのカテゴリ同士に深い関係があり、しかし上下関係を作らないような場合は、カテゴリの本文で関連づけるに留めてください。

図 2: Wikipedia:カテゴリの方針#カテゴリの構造

<sup>2</sup><https://ja.wikipedia.org/wiki/Wikipedia:カテゴリの方針>

割の基準として用いられたカテゴリを組み合わせで命名され、これらの各々のカテゴリを親カテゴリあるいは、先祖のカテゴリとして有する。

カテゴリの包含する記事の数に制限はありませんが、包含する記事が多いカテゴリはしばしばより具体的なサブカテゴリに分割されます。また、既存のカテゴリに対して、それらを一般化した上位カテゴリを作成し、そこに既存のカテゴリをサブカテゴリとして含めることもあります。こうしてできたサブカテゴリのうち、基本的に重なりがなく、上位カテゴリの内容を概ねカバーしている様なサブカテゴリの組のことを、(その上位カテゴリの)「分割として機能しているカテゴリ」と呼びます。

図 3: Wikipedia:カテゴリの方針#分割として機能するカテゴリ

知識工学的な観点、特に、名付き実体の概念分類を与えるために、カテゴリを利用しようとする際に、この分割は、不必要にカテゴリ階層を複雑にするため、その取り扱いについて、考える必要がある。

## 2.2 Wikipedia カテゴリの分析

前小節で述べたように、Wikipedia カテゴリの階層関係は、通常の知識工学的な観点から想定するような親子関係の間に推移的な包含関係が成り立つような階層関係ではなく、利用者の閲覧性の向上を図るための親子関係などが含まれている。

この状況を理解するために、我々は、日本語 Wikipedia のカテゴリ階層を対象として、Wikipedia 特有のメンテナンスのためのカテゴリ (隠しカテゴリのサブクラスに分類され、記事の編集の状況などについての分類などを示すカテゴリ) を除く、全てのカテゴリについて、その記述パターンと、親子関係の分析を行った [14, 13]。

この結果、日本語 Wikipedia の多くのページが「A の B」という形の表記になっていること、多くの場合、B に相当する名詞が分類を表し、A が分割の基準を表していることなどが確認された。また、その分割されたカテゴリ「A の B」の多くが「A」や「B」を親カテゴリに持つこと、また、「A1 の B」→「A2 の B」(A1:アジア→A2:日本)のような、分割の基準を考慮した親子関係や、「A の B1」→「A の B2」(B1:学者→B2:科学者)といった分類の詳細化(クラス・サブクラス関係)が多く存在することが確認された。日本語 Wikipedia オントロジ [2] におけるクラス・サブクラス関係の抽出に、「の」を使ったパターンが用いられているのは、この性質を利用していると考えられる。このようなカテゴリ

が存在することについて、英語版 Wikipedia<sup>3</sup> においては、カテゴリを、分類を表す Set カテゴリ (Cities)、トピックを表す Topic カテゴリ (France) と、その組合わせである Set-and-topic カテゴリ (Cities in France) として分類している。日本語のカテゴリの場合は、上記で述べた「A の B」の多くについては、A が Topic カテゴリ、B が Set カテゴリで、A の B が Set-and-topic カテゴリに分類されると考えられる。

## 2.3 Wikipedia カテゴリ階層構造の問題

前節の図 1 にあるように、Wikipedia カテゴリについては、その成り立ちとしては、概念分類などに有用な情報として付与されることになったものである。しかし、Wikipedia の規模の拡大などにもとない、前述の「分割として機能するカテゴリ」などが導入され、概念分類の詳細化にあたるクラス・サブクラス関係だけではなく、クラス・インスタンスの関係や制約条件の詳細化といった様々な関係が含まれることとなった。

図 4 に典型的な関係を含む Wikipedia カテゴリの階層構造を示す。<sup>4</sup>このように、Wikipedia のカテゴリの親子関係には、様々なタイプの親子関係が混在しており、その全てを親子関係として単純化したカテゴリ階層から情報を用いると、「トヨタ自動車の人物」という「人物」の分割されたカテゴリが「企業」を先祖(上位)のカテゴリに所属するといった不適切な関係を導くこととなる。

また、この結果、本来、禁止されているループ構造を含むカテゴリ階層関係が定義されてしまっている。実際の日本語 Wikipedia<sup>5</sup>のカテゴリ階層において、ループ検出をしたところ、58 件のループが確認された [15]。このループには、「ロンドン」→「ロンドンの地方自治」→「グレーターロンドン」→「ロンドン」といった一つ一つを見ると親子関係が成り立つが、全体としてループになっているようなものが存在することが確認された。

## 3 Wikipedia カテゴリオントロジ

我々は、前節で述べた問題を踏まえ、これらの親子関係を分類し、推移的な包含関係が成り立つ関係とそうでない関係に分類すると共に、ユーザが適切な観点で、Wikipedia カテゴリの階層関係を再構成するための枠組として Wikipedia カテゴリオントロジ (WCO) を提案している [16, 17]。

<sup>3</sup>[https://en.wikipedia.org/wiki/Wikipedia:Category\\_organization](https://en.wikipedia.org/wiki/Wikipedia:Category_organization)

<sup>4</sup>説明を分かりやすくするために、途中の親子関係の一部を簡略化している。

<sup>5</sup>2017 年 10 月 20 日のダンプデータを利用

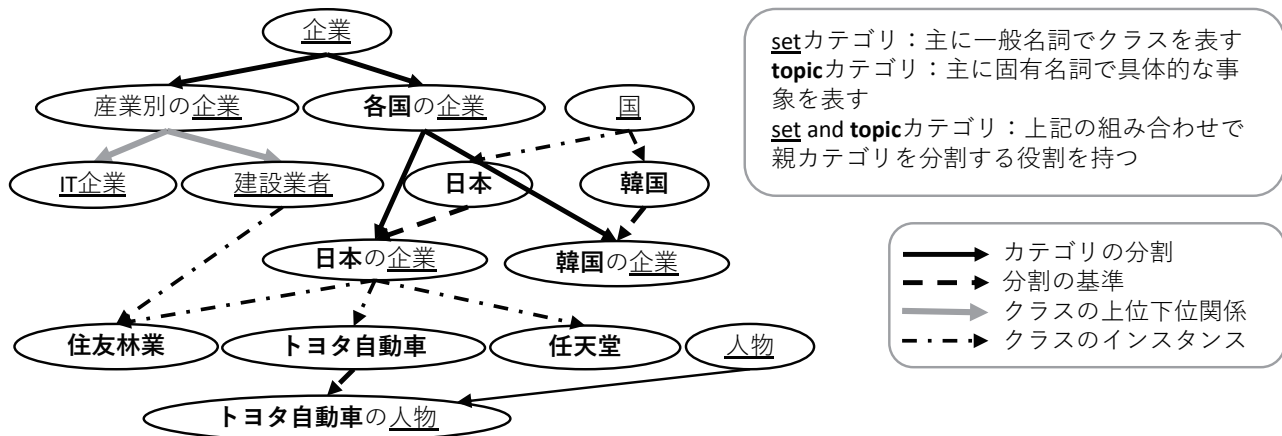


図 4: Wikipedia カテゴリ階層に含まれる様々な関係

この WCO は、カテゴリと、カテゴリ間の親子関係のタイプを定義するための基本語彙を Wikipedia カテゴリオントロジーの Core Vocabulary として定義するとともに、具体的に日本語 Wikipedia の全てのカテゴリ（ただし、先にも述べた Wikipedia 特有のメンテナンスのためのカテゴリを除く）とその関係を分類したデータを作成し、Linked Open Data として公開している (<http://wcontology.org/>)

本節では、この Wikipedia カテゴリオントロジーについて説明すると共に、具体的な利用方法について紹介する。

### 3.1 Wikipedia カテゴリとカテゴリ間の関係の分類

英語版の Wikipedia では、Wikipedia のカテゴリは、主に概念分類を表す Set カテゴリ、トピックを表す Topic カテゴリと、それらの組み合わせである Set-and-topic カテゴリに分類をしていた。しかし、実際にカテゴリを分析すると、「2000 年の日本」のように、Topic と Topic の組み合わせであったり、「京都市の寺院の画像」のように、複数の Set や Topic の組み合わせが存在したりすることが確認された。この分析結果を受けて、WCO では、カテゴリを Set, Topic の各々について、制約付きと制約なしに分類した次の 4 種類に分類することとした。

**Set カテゴリ (setCategory)** 主に分類を表すカテゴリ。

**Topic カテゴリ (topicCategory)** 特定の固有物や事象 (トピック) を表すカテゴリ

**制約付き Set カテゴリ (constrainedSetCategory)**

上記の Set カテゴリから分割され、主に、分割の基準 (制約) を表す言葉と組み合わせられたカテゴリ

**制約付き Topic カテゴリ (constrainedTopicCategory)**

上記の Topic カテゴリから分割され、主に、分割の基準 (制約) を表す言葉と組み合わせられたカテゴリ

また、カテゴリ間の関係については、大きく分けて、推移律が成り立つ関係と、成り立たない関係に分類し、各々について、以下のような小分類を設定した。

#### ● 推移律が成立しない関係

**制約としての利用 (usedForConstraint)** この関係では、親カテゴリが子カテゴリの制約として利用される。例: 「2000 年」 → 「2000 年の日本」、「日本」 → 「日本の大学」

**クラス・インスタンス (instanceOf)** この関係では、子カテゴリが親カテゴリのインスタンスである。例: 「日本の国立大学」 → 「北海道大学」

**Narrower(narrower)** この関係は、上記にあてはまらない推移律が成立しない親子関係である。例: 「演劇」 → 「脚本」、「北海道の大学」 → 「函館大学の教員」<sup>6</sup>

#### ● 推移律が成立する関係

**クラス・サブクラス (subclassOf)** この関係では、子カテゴリが親カテゴリのサブクラスである。制約の有無に関わらず、Set に相当する部分にクラス・サブクラスの関係が成り立つ場合には、この関係を付与する。例:

<sup>6</sup>Wikipedia では、カテゴリに属するページが少ないカテゴリの作成が認められないため、本来なら「北海道の大学」 → 「函館大学」 → 「函館大学の教員」となった方が良いようなカテゴリ間の関係がこのような形で表現される

「俳優」→「女優」、「日本の俳優」→「日本の女優」、「日本の俳優」→「歌舞伎役者」<sup>7</sup>

**制約の付与 (addConstraint)** この関係では、親カテゴリに制約を加えて子カテゴリが作成されている。例: 「人物」→「日本の人物」、「島」、「大陸別の島」

**年代 (age)** この関係では、子カテゴリが親カテゴリの年代の一部である。例: 「20 世紀」→「1980 年代」、「1980 年代」→「1981 年」

**地理的包含関係 (geography)** この関係では、子カテゴリが親カテゴリの地理的な一部である。例: 「東アジア」→「日本」、「2000 年のアジア」→「2000 年の日本」

**制約の詳細化 (specifiedConstraint)** 子カテゴリの制約が、親カテゴリの制約を詳細化 (より制約が強い) したものとなっている関係について、上記の 2 つにあてはまらないもの。例: 「大陸別の島」→「ヨーロッパの島」

**Narrower Transitive(narrowerTransitive)** この関係は、上記にあてはまらない推移律が成立する親子関係である。例: 「20 世紀のヨーロッパ」→「ドイツ帝国」、「2013 年のテニス」→「2013 年の ATP ワールドツアー」

推移律が成立する関係のうち制約の詳細化については、典型的事例である年代と地理的包含関係を別の分類としている。

これらの語彙を以下のように定義した。DBpedia において、カテゴリ間の関係を、親が主語で子供が述語として、skos:broader で定義している。これに対し、WCO では、子供を主語で親が述語となるので、全ての関係を skos:narrower のサブプロパティとして定義し、さらに、推移律が成り立つカテゴリについては、skos:narrowerTransitive のサブプロパティとして定義した。

### 3.2 日本語版 Wikipedia カテゴリの分類

次に、ここで提案した分類に基づいて、実際の Wikipedia カテゴリ<sup>8</sup>を分類し、Linked Open Data として公開した。表 1 に各分類に分類されたカテゴリの数を示す。この数値から分かるように、複数のカテゴリの組合わせで表される制約付きの constrainedSetCategory や constrainedTopicCategory が、約 2/3 を占めており、分割のためのカテゴリによりカテゴリ数が増えていることが確認できる。

<sup>7</sup>歌舞伎役者は日本以外にいないので、「日本の」という制約が明示的に必要ない

<sup>8</sup>2017 年 10 月 20 日のダンプデータを利用

表 1: タイプ後とのカテゴリ数

カテゴリタイプ	Number of categories
<i>setCategory</i>	10,748
<i>topicCategory</i>	44,525
<i>constrainedSetCategory</i>	117,994
<i>constrainedTopicCategory</i>	10,333
Total	183,600

また、親子関係については、表 2 に分類結果を示す。これを見ると、最も多い関係は、specifiedConstraint である。また、subclassOf についても、制約付きのカテゴリ間の関係につけられている場合も多くカテゴリ数としても多くを占める制約付きのカテゴリ間の関係がかなりの部分を占めることが分かる。

表 2: Number of subcategory relationships for each type

親子関係のタイプ	親子関係の数
<i>usedForConstraint</i>	41,885
<i>instanceOf</i>	64,319
<i>narrower</i>	98,201
<i>subclassOf</i>	85,965
<i>age</i>	3,172
<i>geography</i>	8,244
<i>addConstraint</i>	8,301
<i>specifiedConstraint</i>	133,579
<i>narrowerTransitive</i>	7,408
Total	451,074

### 3.3 WCO を用いた分析事例

WCO を用いることにより、先に紹介したカテゴリの親子関係を制限なくたどることにより発生する不適切な概念の包含関係の問題や、ループの問題が、どのように解決されるかについて説明する。

まず、「日本の企業」→「トヨタ自動車」→「トヨタ自動車の人物」では、最初の関係が instanceOf、次の関係が usedForConstraint であり、どちらも推移律が成り立たないため、「人物」の分割されたカテゴリが「企業」の分割されたカテゴリに包含されるという解釈にはならないことが示される。また、「ロンドン」→「ロンドンの地方自治」→「グレーターロンドン」→「ロンドン」のループについて考える。この時、3 つの関係は addConstraint、instanceOf、narrower に分類され、それぞれ、推移律が成り立たない。そのため、このようなループをたどること辞退が不適切と判断することができ、問題に応じた適切なカテゴリの親子関係のみをループを含まない形で切り出すことが可能となる。

これ以外にも、分割されたカテゴリを、分割前のカテゴリに戻すための SPARQL について考える。これ

```
@prefix wcoc: <http://wcontology.org/core#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
wcoc:topicCategory rdfs:subClassOf wcoc:category .
wcoc:setCategory rdfs:subClassOf wcoc:category .
wcoc:constrainedSetCategory rdfs:subClassOf wcoc:setCategory .
wcoc:constrainedTopicCategory rdfs:subClassOf wcoc:topicCategory .
wcoc:narrower rdfs:subPropertyOf skos:narrower .
wcoc:usedForConstraint rdfs:subPropertyOf wcoc:narrower .
wcoc:instanceOf rdfs:subPropertyOf wcoc:narrower .
wcoc:narrowerTransitive rdfs:subPropertyOf wcoc:narrower .
wcoc:narrowerTransitive rdfs:subPropertyOf skos:narrowerTransitive .
wcoc:age rdfs:subPropertyOf wcoc:narrowerTransitive .
wcoc:geography rdfs:subPropertyOf wcoc:narrowerTransitive .
wcoc:addConstraint rdfs:subPropertyOf wcoc:narrowerTransitive .
wcoc:specifiedConstraint rdfs:subPropertyOf wcoc:narrowerTransitive .
wcoc:subclassOf rdfs:subPropertyOf wcoc:narrowerTransitive .
```

図 5: Definition of the core vocabulary used by WCO

は、与えられたカテゴリから推移律が成り立つ関係だけをたどって得られる全てのカテゴリが分割後のカテゴリに相当するため、図6のようなクエリとなる<sup>9</sup>。この例の場合は、「日本の大学」の分割されたカテゴリとして、「日本の公立大学」といった組織母体に関する分類や「北海道の大学」といった地理的な分割がされたカテゴリなど71のカテゴリが得られた。

さらに、このカテゴリ群に属するインスタンスに相当するカテゴリを抽出するクエリが図7となる。この結果として、「和歌山大学」、「近畿大学」といった490の大学のカテゴリ(一部、大学院なども含む)を得られることとなった。

また、概念階層に関する情報に注目すると、各概念のサブクラスに相当するカテゴリを抽出する場合にも、同じく、推移律が成り立つ関係だけをたどって得られる全てのカテゴリの内、setCategoryのものに限定することで得ることができるため、図8のようなクエリとなる。この例の場合は、「人物」のサブクラスとして、「人道支援家」、「タタール人」といった3923のカテゴリが存在することが確認された。ただ、現時点では、ConstrainedSetとSetを区別する際に「AのB」のように助詞でつながれていることを区別の条件にしていたため、「在アイスランド大使」や、「在アイルランド大使」のように、国が違うだけの大使をそれぞれ個別のSetとして数え上げている。こちらについては、今後、その扱いについてさらなる検討が必要であると考

<sup>9</sup> これらのクエリについては、<http://wcontology.org/> の endpoint で実行可能である。

えている。

このように、WCOを用いることで、Wikipediaカテゴリ階層から、知識工学的な観点で必要となる情報を取り出し利用できることが確認できた。

## 4 おわりに

本論文では、我々がこれまでに行ってきたWikipediaカテゴリに関する分類の研究を踏まえて作成したWikipediaカテゴリオントロジーについて紹介した。このWikipediaカテゴリオントロジーを用いることにより、これまで体系的に活用されてこなかったWikipediaカテゴリの階層構造を選択的に利用することが可能となる。今後の課題としては、本オントロジーを発展するWikipediaに適用するためには、現在、手作業で作成したカテゴリやカテゴリ間の分類について、機械学習などを用いることにより継続的にその内容を更新可能な枠組へと進化させる必要がある。また、その成果を用いた他言語への展開も同じく研究課題である。

## 謝辞

本研究の成果をまとめるにあたり北海道大学大学院情報科学研究院の知識ベース研究室の学生、中川君、小坂橋君など多くの学生に協力を頂いた。また、本研究の一部は、科研費基盤研究(B)18H03338により行われた。ここに記して、謝意をあらわす。

```

PREFIX wcoc:<http://wcontology.org/core#>
PREFIX jwcor: <http://ja.wcontology.org/resource/Category:>

select distinct ?jcc where {
  ?jcc
  (wcoc:subClassOf | wcoc:addConstraint | wcoc:specifiedConstraint
  | wcoc:narrowerTransitive | wcoc:geography |wcoc:age )+ jwcor:日本の大学 .
}

```

図 6: 日本の大学に関する分割されたカテゴリ群の獲得

```

PREFIX wcoc:<http://wcontology.org/core#>
PREFIX jwcor: <http://ja.wcontology.org/resource/Category:>

select distinct ?s where {
  ?jcc
  (wcoc:subClassOf | wcoc:addConstraint | wcoc:specifiedConstraint |
  wcoc:narrowerTransitive | wcoc:geography |wcoc:age )+ jwcor:日本の大学 .
  ?s wcoc:instanceOf ?jcc .
}

```

図 7: 日本の大学に関する分割されたカテゴリ群に属するインスタンスカテゴリ

```

PREFIX rdf:<http://www.w3.org/1992/02/22-rdf-syntax-ns#>
PREFIX wcoc: <http://wcontology.org/core#>
PREFIX jwcor:<http://ja.wcontology.org/resource/Category:>

select distinct * where {?s
  (wcoc:subClassOf | wcoc:addConstraint | wcoc:specifiedConstraint |
  wcoc:narrowerTransitive |wcoc:geography |wcoc:age )+ jwcor:人物 .
  ?s rdf:type wcoc:setCategory}

```

図 8: 「人物」に関するサブカテゴリで setCategory に分類されるものの一覧の獲得

## 参考文献

- [1] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Soren Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. DBpedia - a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 7, No. 3, pp. 154 – 165, 2009.
- [2] 玉川奨, 桜井慎弥, 手島拓也, 森田武史, 和泉憲明, 山口高平. 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, Vol. 25, No. 5, pp. 623–636, 2010.
- [3] Thomas Pellissier Tanon, Gerhard Weikum, and Fabian Suchanek. YAGO 4: A reason-able knowledge base. In Andreas Harth, Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo, Heiko Paulheim, Anisa Rula, Anna Lisa Gentile, Peter Haase, and Michael Cochez, editors, *The Semantic Web*, pp. 583–596, Cham, 2020. Springer International Publishing.
- [4] Torsten Zesch and Iryna Gurevych. Analysis of the Wikipedia category graph for NLP applications. In *Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing*, pp. 1–8, Rochester, NY, USA, 2007. Association for Computational Linguistics.
- [5] Simone Paolo Ponzetto and Michael Strube. Knowledge derived from wikipedia for computing semantic relatedness. *Journal of Artificial Intelligence Research*, Vol. 30, No. 1, pp. 181–212, October 2007.
- [6] Mohamed Ali Hadj Taieb, Mohamed Ben Aouicha, and Abdelmajid Ben Hamadou. Computing semantic relatedness using wikipedia features. *Knowledge-Based Systems*, Vol. 50, No. 0, pp. 260 – 278, 2013.
- [7] Peng Yan and Wei Jin. Mining semantic relationships between concepts across documents incorporating wikipedia knowledge. In Petra Perner, editor, *Advances in Data Mining. Applications and Theoretical Aspects*, Vol. 7987 of *Lecture Notes in Computer Science*, pp. 70–84. Springer Berlin Heidelberg, 2013.
- [8] Muhammad Jawad Hussain, Shahbaz Hassan Wasti, Guangjian Huang, Lina Wei, Yuncheng Jiang, and Yong Tang. An approach for measuring semantic similarity between wikipedia concepts using multiple inheritances. *Information Processing & Management*, Vol. 57, No. 3, p. 102188, 2020.
- [9] Paolo Boldi and Corrado Monti. Cleansing wikipedia categories using centrality. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW ’16 Companion*, p. 969–974, Republic and Canton of Geneva, CHE, 2016. International World Wide Web Conferences Steering Committee.
- [10] Nicolas Heist and Heiko Paulheim. Uncovering the semantics of wikipedia categories. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, *The Semantic Web – ISWC 2019*, pp. 219–236, Cham, 2019. Springer International Publishing.
- [11] Nicolas Aspert, Volodymyr Miz, Benjamin Ricaud, and Pierre Vanderghenst. A graph-structured dataset for wikipedia research. In *Companion Proceedings of The 2019 World Wide Web Conference, WWW ’19*, p. 1188–1193, New York, NY, USA, 2019. Association for Computing Machinery.
- [12] 吉岡真治. Wikipedia のカテゴリ階層関係の分類を用いた日本語 wikipedia オントロジーの分析. 2014 年度人工知能学会全国大会 (第 28 回) 論文集, 2014. CD-ROM 2J3-4.
- [13] Masaharu Yoshioka. Analysis of japanese wikipedia category for constructing wikipedia ontology and semantic similarity measure. In *Information Retrieval Technology 10th Asia Information Retrieval Societies Conference, AIRS 2014, Kuching, Malaysia, December 3-5, 2014 Proceedings*, pp. 470–481. Springer-Verlag GmbH, 2014. LNCS8870.
- [14] 藤原嵩大, 吉岡真治. Wikipedia の階層関係を分析するためのカテゴリパターン の提案. 2012 年度人工知能学会全国大会 (第 26 回) 論文集, 2012. CD-ROM 2C1-NFC2-4.
- [15] 中川嵩教, 小坂橋佳晃, 吉岡真治. カテゴリの親子関係の種類に基づく wikipedia カテゴリの再整理.



2019 年度人工知能学会全国大会 (第 33 回) 論文集,  
2019. CD-ROM 1K2-J-4-05.

- [16] 中川嵩教, 小板橋佳晃, 吉岡真治. Wikipedia カテゴリオントロジーを利用した wikipedia ページのクラス分類. 2020 年度人工知能学会全国大会 (第 34 回) 論文集, 2020. CD-ROM 4H3-GS-4-05.

- [17] Masaharu Yoshioka and Takanori Nakagawa. Wikipedia category ontology: A framework for utilization of the wikipedia category structure by knowledge engineers. In *Proceedings of the ISWC 2020 Satellite Tracks (Posters & Demonstrations, Industry, and Outrageous Ideas) co-located with 19th International Semantic Web Conference (ISWC 2020)*, 2020.