

# ニコニコ動画で視聴をより楽しむためのアプリケーションのためのニコニコ用語辞書拡張

川本 洋輔<sup>1,a)</sup> 西山 泰貴<sup>1,b)</sup> 小林 暁雄<sup>1,c)</sup> 坂地 泰紀<sup>2</sup> 増山 繁<sup>1,d)</sup>

概要：ニコニコ動画のコメントは基本的に画面上を流れているだけであり、視聴を楽しむには物足りない。そこで、コメントに関連するタグや大百科記事などの情報を自動で付加する、コメントから特定の嗜好の利用者を発見するなどの様々な機能があれば動画の視聴をより楽しむことができる。そのために我々はニコニコ大百科から、コメントに出てくる用語の機械可読な拡張固有表現辞書を構築し、コメントに様々な情報を付加することを可能とする。

## NicoNico Buzz Phrase Dictionary for Enjoying NicoNico Movies

KAWAMOTO YOUSUKE<sup>1,a)</sup> NISHIYAMA YASUKI<sup>1,b)</sup> KOBAYASHI AKIO<sup>1,c)</sup> SAKAJI HIROKI<sup>2</sup>  
MASUYAMA SHIGERU<sup>1,d)</sup>

**Abstract:** Comments on the movie in NicoNico Douga are just shown without any interaction. It might be more enjoyable that these comments have some functions such as: users can get information of named entity on the shown comment by the graphical interaction, or shows any user groups who are interested in some kind of objects in the movie by their posted comments. Thus, we provide a NicoNico buzz phrase dictionary that includes Extended Named Entity(NE) information for realizing those applications.

### 1. はじめに

ニコニコ動画は、株式会社角川ドワンゴが提供する動画共有サイトであり、利用者数、投稿動画数について、日本国内のサービスとしては、最大の規模を誇る<sup>\*1</sup>。ニコニコ動画では、投稿された動画に対して、視聴者や動画投稿者（以下、ユーザ）が動画上の特定のタイミングで表示されるコメントを投稿することができる。この機能により、ユーザ同士の間でコミュニケーションを行うことができる。各動画はユーザの望むときに再生することができるため、コ

メントによるコミュニケーションは非同期に行われている。このため、各ユーザは動画の視聴やコメント投稿のために特定の時間を拘束されるようなことはない<sup>\*2</sup>。その手軽さもあって、ニコニコ動画では、コメント投稿とそれによるコミュニケーションが積極的に行われている。多いものでは、1 動画に対して 1,000 万件以上のコメントが投稿されることもあり、コメント機能が今日のニコニコ動画の発展の大きな要因となっていると考えられる。

このため、各動画に対して投稿されるコメントには、ユーザの動画に対する感想だけでなく、主張の対立や同意などのコミュニケーションから発生する議論と、それに伴うユーザの感情や興味嗜好なども反映されている。

このような特徴から、コメントデータをユーザの行動分析や嗜好分析などに利用し、動画を推薦する研究 [16], [17] や、コメントの他ユーザにとっての有用性を評価する研究 [2], [12] も行われている。ところが、コメントデータは、動画の任意の時間・位置に動画上に配置されるため、ユー

<sup>1</sup> 豊橋技術科学大学  
Toyohashi University of Technology

<sup>2</sup> 成蹊大学  
Seikei University

a) kawamoto@la.cs.tut.ac.jp

b) nishiyama@la.cs.tut.ac.jp

c) a-kobayashi@cs.tut.ac.jp

d) masuyama@tut.jp

\*1 投稿動画数が 1,200 万件超、総再生数 628 億回超、総コメント数 47 億超 (2015 年 7 月現在)。ユーザ数は登録ユーザ数が 3936 万人、有料会員登録数が 223 万人 (2014 年 3 月)

\*2 特定時間にのみ放送されるニコニコ生放送を除く



図 1 ニコニコ動画ページ例

Fig. 1 Example of the movie page on NicoNico

ザの動画に対する即時的な感情や特定シーンに限定された感想によるものが多く、非常に短い文字列で表記されることが多い。このため、コメントのうち、各研究それぞれにとって有用なもののみを判別することは、非常に困難となっている。また、SNSとして、大多数のユーザが積極的にコミュニケーションを行っていることから、独自の言語現象も発生していることが観測されている [5]。

このため、これらのコメント特有の表現を一覧する辞書が存在すれば、ニコニコ動画のコメントを利用する応用研究におけるノイズ除去などの処理の精度向上が期待できる。

そこで、本研究ではコメント中の表現の中でも、頻繁に用いられている、固有名詞や定型句について、品詞情報に加え、意味分類を表す固有表現情報を付与した辞書の構築を行った。

## 2. ニコニコ動画の視聴ページ構造

ニコニコ動画における投稿された動画のウェブページ（以降、動画ページ）と、動画ページ中の各機能について解説する。動画ページの例を図 1 に示す。

動画ページには、投稿者によって付与された、動画タイトル、動画概要、動画自体と、視聴者によって付与されるコメントが動画上、あるいは、動画右側のコメント一覧に表示される。

また、投稿者・視聴者両方が付与、編集できる構造として、動画を分類するためのタグと動画に関連する商品（ニコニコ市場）が存在する。特に、タグは編集者が自由に記述できるため、動画を階層構造などで明確に分類する構造

というよりは、Folksonomy として機能している。ただし、タグが設定できる上限が 1 動画につき 10 件に設定されているため、一般的な folksonomy よりも、洗練される傾向にある [14]。

## 3. 関連研究

ニコニコ動画では、動画を見た視聴者が、その動画に触発されて新たな動画を作成、投稿するといった、共創活動が頻繁に行われている。このため、この共創活動に関する動画の関連性を見出す、あるいは、予測するといった研究が行われている [10], [11], [15]。

浜崎ら [15] は、動画概要中の他動画へのリンク構造や、カテゴリとなっているタグの分類情報などから動画間関係を示すネットワークグラフを作成し、多角的にユーザ同士のコミュニケーションによる創造活動の分析を行っている。しかしながら、最もユーザ間のコミュニケーションが明確に出現するコメントについては言及されていない。

伊藤ら [11] では、視聴者の興味を示すタグを発見することで、動画の価値を判定できることを示唆している。しかし、最も視聴者の興味が直接書き込まれる、コメントについては対象としていない。

大田ら [10] は、タグを用途別に 3 種類に分類し、それらを特定ジャンルが成形される様子を観測するための情報として使用できる可能性があることを示している。特に、視聴者の動画に対する感想を反映しているタグは、今後新たな動画ジャンルの発生を示す可能性がある他、コメントがこの種のタグとして取り上げられることがあるため、コメントの解析の必要性について言及している。

これらの研究では、最もユーザの興味・嗜好や、共創活動につながるコミュニケーションが反映されているコメントデータが対象とされていない。現状のコメントデータは、後述のコメントデータを使用した研究でも報告されているように、ノイズ除去が困難である。このため、構造的に利用しやすいタグ等と比較して、整備に多大なコストがかかり、研究対象として使用しづらいことが大きな要因である可能性がある。

このため、コメントデータの自動分類、あるいは、自動分類のための指標となるデータの構築は、ニコニコ動画を通じたユーザユーザ間のコミュニケーションをより詳細に解析する上で、多大な価値があると考えられる。

コメントデータを使用した研究としては、[3], [6] が存在する。Wu ら [3] は、動画を単位時間（シーン）に区切り、異なるシーンに投稿されたコメント間の対応付けを行うことで、コメント投稿者間の対話の抽出を試みている。Wu らは、ノイズ除去や単語分割において大きな課題が残っていると報告している。

池田ら [6] は、コメントデータにアノテーションを行うことで、コメントが動画ページ中のどの部分に対して言及し

ているのかを明確にすることにより、コメント解析精度向上につながるデータの構築を試みている。池田らの結果から、動画に付与されたコメントの殆どは動画内の特定シーンに結びつくことから、動画内に使用されているコンテンツが、コメント内に含まれていることを判定する必要があるとしている。

このように、これらの研究では、コメント内に出現する単語の判定が精度向上に不可欠であると考察されている。本研究によって構築される辞書を用いることによって、これらの応用研究の精度が向上できると期待される。

また、ニコニコ動画コメントにおける言語現象を対象とした研究として、内山ら [5] が行われており、2ちゃんねるにおける言語現象と同様の発生メカニズムに基づいていることが報告されている。

このことから、2ちゃんねるを対象とした言語解析の研究 [4], [13] の成果がニコニコ動画に対しても適用できる可能性がある。これらの研究では、形態素解析のための辞書構築や、2ちゃんねるに特徴的な表現の辞書を構築しており、本研究の成果である辞書を用いることで、このような応用研究がニコニコ動画のコメントデータに対しても可能になると考えられる。

このため、ニコニコ動画コメントを対象としてこのような応用研究を行うには、固有名詞や定型句の辞書は必要不可欠である。しかしながら、ニコニコ動画コメント解析のための辞書構築は、我々の知る限り、現在行われていない。よって、本研究では、ニコニコ動画固有の単語、定型句の一覧辞書の構築を行うことを目的とする。

#### 4. 辞書構築について

本研究では、ニコニコ動画コメントに使用される単語や表現を収録した辞書の構築を行う。しかしながら、ニコニコ動画に出現する全コメントの収集とその品詞情報を動画を見ながら付与することは、コメント数・動画数が莫大な数に上るため、非常に困難である。このため、本研究では、コメントに出現し得る言葉が多数収録された、ニコニコ大百科(ニコニコ動画視聴者によって構築されたウェブ百科事典。詳細は後述)より、辞書に収録する対象となる語を取得、品詞情報の付与を行う。

この際、前章で述べたような関連研究に用いるうえで、IPAdic\*3 などで定義された品詞体系を用いる場合、ほとんどの語彙が固有名詞(一般)となってしまうと考えられる。

そこで、IPAdic 準拠の品詞、品詞細分類に加えて、単語・定型句の上位概念を拡張固有表現として付与した固有表現辞書を構築する。例えば、ボーカロイドキャラクターの初音ミクであれば、“初音ミク、固有名詞、人名、ボーカロイド”といったような、品詞細分類に加えて、対象用語、ま

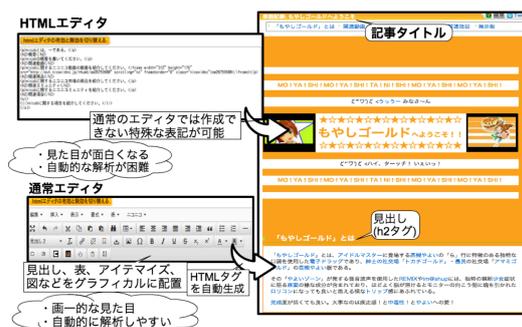


図 2 ニコニコ大百科記事ページ例

Fig. 2 An Example of the article of NicoNico encyclopedia

たは、表現の上位概念をさらなる分類として付与する。

#### 4.1 ニコニコ大百科

ニコニコ大百科は、ニコニコ動画、ニコニコ静画などの、ニコニコに関する各種サービスに関連した事柄の詳細を記事として収録したウェブ上の大百科辞典である。収録記事は利用者によって日々作成・修正が行われており、現在、15万件以上の記事が存在する。

図 2 に大百科記事“もやしゴールド”の例を示す。このように、ウェブ上の大百科辞典であるウィキペディアのように、記事が記載されており、ウィキペディアからの固有表現抽出を行っている研究 [1], [7], [8], [9] の成果が使用できる可能性がある。しかしながら、例のように、記事冒頭に記事解説とは関係のない文言・画像などが使用されるなど、ニコニコ大百科の記事は、ジョーク・パロディなどの側面が強く、文章も定型的に書かれていない場合も存在する。

このため、本研究では、[9] らのように、特定概念に属する固有表現がリストアップされている、一覧記事をもちいて、そこに一覧されている固有表現を抽出し、記事タイトルを各固有表現の上位概念として付与することで、高精度かつ、多くの用語の取得を行う。

ニコニコ大百科では、記事の記述は、Wiki 記法のような独自の方式によって作成される。このため、殆どの一覧記事では、この記法に則った、表形式、アイテムズ形式、箇条書きによって記載されている。

そこで、本研究では、これらの記事構造を解析するための幾つかのパターンを用意し、これを各記事に適合させることによって、ニコニコ特有の用語の抽出を行う。

#### 4.2 固有表現及び定型句抽出手法

ニコニコ大百科から固有表現を抽出するにあたり、ニコニコ大百科の一覧記事から用語を抽出する。その際に、一覧記事を取得する必要がある。ニコニコ大百科では、一覧の一覧という、一覧記事を一覧するための記事が存在するため、ここに含まれる記事を対象として、抽出手法を適用した。提案手法のフローチャートを図 3 に示す。

\*3 形態素解析器 ChaSen 用辞書, <http://osdn.jp/projects/ipadic/>

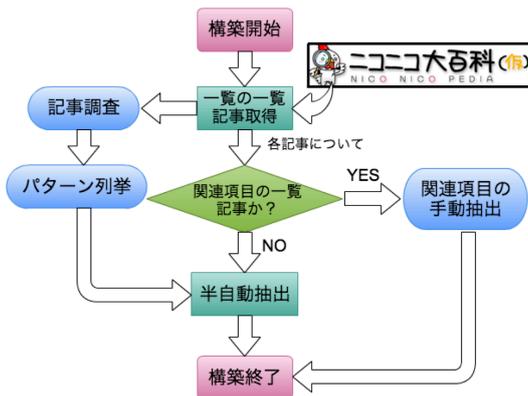


図 3 提案手法フローチャート  
Fig. 3 Flow Chart of the proposed method

図 3 に示すように、まず一覧の一覧記事内に存在する全記事に対し人手により、記述形式の調査と大まかな適用手法の分別を行った。調査結果を用いて、記事の大まかな記載種別毎に 5 種類のパターンとノイズフィルタを構築し、これらを適用する半自動の用語抽出を行った。

また、各記事に半自動抽出手法を適用する前に、収録された概念自体に統一性がない、関連項目の一覧記事をあらかじめリストアップし、別途人手による用語抽出（以降、関連項目抽出手法と略記）を行った。

#### 4.2.1 半自動抽出

ニコニコ大百科の記事ページは、Wiki 記法のような独自記法によってユーザが記載した記事について、システム側が自動で解釈、HTML タグ付与を行っている。このため、付与されている HTML タグはある程度記事間で統一されており、また、本研究で対象としている記事が全て一覧記事であることから、多くの記事が特定のパターンで記載されていることが予備調査から判明している。

このため、特定の HTML タグのパターンと、用語以外の情報の典型的なパターンに対応したパターンマッチングを行うことで、一定数の記事に対して用語抽出とノイズ除去を行うことができる。また、一覧記事は基本的には特定の概念（記事タイトル）に対する下位概念を表す用語を列挙しているため、列挙されている用語は全て同一の品詞のものである。このことから、用語抽出パターンを適用する際に、適用対象の記事のタイトルを人間が確認することで、品詞情報も半自動的に付与することができる。

表 1 に用語抽出のための HTML パターン、表 2 に補足事項等の用語以外の情報を除外するためのパターンの一覧を示す。また、これらのパターンを適用するにあたって、空白文字の半角スペースへの正規化を行った。

表 1 中の表リスト、表、リストはそれぞれ、HTML 表記における表、リスト、または、その組み合わせを用いて用語を一覧している。また、用語定義のための HTML タグである、dt タグを用いて用語を記載しているデータも

表 1 用語抽出のための HTML パターン  
Table 1 Patterns for Phrase Extraction

表記種別	判別に用いたタグ
表リスト	table タグ以下のリスト形式 (li)
表	table タグ以下の表要素 (tr,td)
リスト	リスト形式 (ul) タグ以下の li タグ
用語定義	dt タグに用語、dd タグに用語の解説
用語列挙	dd タグに用語の一覧、dt タグに用語の種別を記載

表 2 ノイズ除去パターン  
Table 2 Patterns for Noise Filtering

ノイズ出現パターン	記述例
丸・角カッコ書き注釈	トロ (Vita 版のみ)
ノイズ文字列	MAD/MAD-PV [動画] [動画]
説明文+コロン”:	第一ドール:水銀燈 (肩書き:人名)
コロン+ノイズ文字列	今治市:東予地方で人口が一番多い
ノイズ文字列	道産子ホイホイ 道民ホイホイ

散見されたので、これに対応するパターン”用語定義”も用いた。

これに対して、dt タグに対して、用語の解説を行う dd タグ部分に用語の列挙を記載し、dt タグにはその種別や上位語を記載するケースが存在した。例えば、テレビゲームの一種である、”ロックマン X のキャラクター一覧”という記事では、dt タグに”大ボス”、dd タグ以下に各キャラクターの名前が記載されている。このため、dt タグから用語を取得するパターンと、dd タグから用語を取得するパターンを別々に用意し、適用した。

また、スラッシュ (”/”) 区切りで用語を列挙しているケースが、dd タグで多く見られたため、表 2 のノイズ除去パターンを適用したのち、スラッシュで文字列を分割し、分割後の文字列をそれぞれ対象用語として品詞情報・固有表現情報の付与を行った。

特殊なケースとして、表形式で用語を列挙する際に、表中の特定の枠内に該当する用語が存在しない場合、ハイフン”-”、または、長音記号”ー”の連続でその枠が空であることを示すケースが散見された。このため、このようなハイフンまたは長音記号の連続の除去を、表 2 のノイズ除去後に行った（例えば、MikuMiku ダンス動画という種類の動画投稿者の一覧記事では、索引別の枠を持った表形式を使用しているが、特定の索引に該当する投稿者名が存在しない場合に、”ー”という長音記号の 3 連続で表記を行っている）。

コロン”:”の前後のどちらをノイズとして除去するかについては、現在は対象記事の品詞情報を人手で与える際に、人手により判断を行っている。しかしながら、左右で記事へのリンクが多いか否か、あるいは、文字数が多いか否かといった情報を用いることで自動化することが可能であると考えられる。

表 3 関連項目抽出抽出

Table 3 Rules of Manual Phrase Extraction from Relational Article Pages

付与品詞	判断基準
固有名詞(人名)	映像作品記事のスタッフなどの関連人物
固有名詞(組織)	映像作品記事の制作会社, 登場組織など
固有名詞(地域)	映像作品記事における登場地域など
固有名詞(一般)	上記以外の関連項目
定型句	名言・セリフなど, 固有名詞ではない表現

#### 4.2.2 関連項目抽出手法

”(作品名など)の関連項目一覧”などの, 一覧記事内において, 記事名に対して様々な関連性にある用語がリストアップされている記事については, 自動で用語抽出を行うことが困難であるため, 人手による用語抽出と品詞・固有表現情報付与による辞書登録を行った。この際に, 表 3 に示す基準を設定し, これに基づいて作業を行った。

付与される固有表現情報については, 適切な分類を人手で判断することが困難な記事も存在したため, 一律に”(作品名など)の関連項目一覧”における, ”(作品名など)”の部分进行分类として指定した。

## 5. 実験

実験を行うにあたり, ニコニコ大百科”一覧の一覧”記事と, そこに一覧されている各一覧記事ページを取得した(2015年7月7日)。記事数は全 1,239 記事(一覧の一覧内での遷移を行うアンカーと同一記事へのリダイレクトを除いた有効リンク数)であり, そのうち, 提案手法の半自動抽出の対象となった記事は 575 件, 関連項目抽出の対象である, 関連項目一覧記事は 75 件であった。残りの記事については, ニコニコ大百科の記事作成に関わる記事, 動画の関連商品のリンクの一覧記事, ニコニココミュニティ(ドワンゴが提供しているサービスの一つであり, ニコニコ関連のサービスの利用者が作る SNS グループ)名の一覧記事, 特定 ID の動画一覧, 特定期間に動画ランキングの高かった動画一覧などであり, 今回の用語抽出の対象としてふさわしくないと判断し, 除外した。

このうち, 実際に抽出を行った件数は, 半自動抽出が 541 件, 関連項目抽出が 75 件全件である。半自動抽出対象中で実際に抽出ができなかった記事は, ユーザの HTML 記述ミスによる HTML 解析エラーが発生した記事である。

この結果, 各手法を適用した際の抽出された用語数は, 半自動抽出が 107,121 語, 関連項目抽出が 14,238 語である。

また, これらの抽出結果から, 半自動, または, 関連項目抽出の結果を区別せずに 1,000 件ランダムに抽出した用語について, 人手により正誤判定(抽出と品詞付与が同時に行われるため, 抽出精度と品詞付与の精度を同時に判定)し, 評価者全員が正解と判断した正解率と評価者間の一致率(評価者は工学部学生・教員 3 名)を求めた。その結果

表 4 実験結果

Table 4 Result of Experiment

評価データ	尺度	値
全て	正解率	0.810
	平均 $\kappa$ 値	0.323
半自動のみ	正解率	0.860
	平均 $\kappa$ 値	0.298
関連項目抽出のみ	正解率	0.366
	平均 $\kappa$ 値	0.228

を表 4 に示す。なお, サンプル中, 半自動によるものは 899 件, 関連項目抽出によるものは 101 件であった。

表 4 より, 評価データ中の約 8 割のデータに対して, 全評価者が正解であると判定している。品詞付与については, 半自動については, 各記事がどのような品詞を一覧しているのかを人手で判断して付与, 関連項目一覧記事については, 一語ずつ人手で付与しているため, ほぼ正解であると考えられる。各抽出文字列に対する固有表現の抽出は, 半自動抽出手法については, 抽出した一覧記事タイトルが一覧している項目の上位概念であることが大多数であることから, ほぼ正解であると考えられる。一方で, 関連項目抽出については, 一律に記事タイトルから抽出しているため, 関連性が消失してしまっており, 正解率が著しく低くなっている。

また, 正解率の高さに対し, 評価データに対する  $\kappa$  係数から, 不正解の判定について, 評価者間の揺れが大きいことがわかる。このことから, 評価者によっては判断が困難となるような, 専門の知識を必要とする用語も数多く収録されていると考えられる。

## 6. 考察

### 6.1 エラー解析

エラーを調査したところ, 半自動手法と関連項目抽出手法について, それぞれ異なった原因に基づいたエラーが発生していた。まず, 半自動分類手法についてのエラーと原因の組みを表 6 に示す。

半自動分類手法では, 表 6 の”パターン一致誤り”のように, ノイズ除去パターン漏れが多く発生しており, 新たなパターンを用意することでとりのぞけるエラーが多かったため, 今後改善できると考えられる(エラー例では, 本来抽出すべき用語の後に記述されている, 用語の解説部分を誤って抽出している)。また, HTML タグからの抽出パターンでも漏れが発生しており, パターンの改善を行う必要がある。

一方で, 品詞細分類付与の誤りも発生しており, 表中の, ”HY”は芸能人名であるため, IPAdic 同様の”人名”と付与されているが, 実際にはミュージシャンのグループであるため, 評価者によっては”組織”であると判断されており,

表 5 半自動抽出のエラー例

Table 5 Errors of Semi Automatic Phrase Extraction

パターン一致誤り	
抽出した用語	荒れ果てた地上を再生するために過去の遺跡や施設を調査する非営利団体。
付与品詞	固有名詞（組織）
固有表現	アーマード・コアシリーズに登場する企業・組織
付与品詞誤り	
抽出した用語	HY
付与品詞	固有名詞（人名）
固有表現	ミュージシャン
付与固有表現情報誤り	
抽出した用語	アレキサンドリア
付与品詞	固有名詞（一般）
固有表現	ガンダムシリーズのMS・MAの一覧

表 6 関連項目抽出のエラー例

Table 6 Errors of Manual Phrase Extraction

品詞情報付与誤り	
抽出した用語	増田順一
付与品詞	固有名詞（人名）
固有表現	ポケモン
品詞情報付与誤り	
抽出した用語	タルコンガ
付与品詞	固有名詞（一般）
固有表現	ドンキーコング

人間でも評価が難しいと考えられる。その一方で、“HY”がミュージシャンであることには疑いがないため、本手法により付与された固有表現の分類は正しいと考えられる。このため、応用研究によっては、この種のエラーは問題にはならないと考えられる。

また、一覧記事自体の誤りもごく少数ではあるが存在していた。“ガンダムシリーズのMS・MAの一覧”記事に記載されていた、“アレキサンドリア”は、記事自体は特定のアニメ作品シリーズに登場するロボットを一覧するためのものであるが、ロボットの配備されている船舶も同時に記載されていたため、エラーとなっている。この種のエラーを自動で判定することは難しいと考えられるほか、評価者についても、このアニメ作品シリーズに詳しくなければ、正しく評価できないようなエラーとなっている。このため、この種のエラーの修正には、各一覧記事毎に、その分野に精通した評価者が必要となると考えられる。

次に、関連項目抽出手法のエラー例を表 7 に示す。

こちらは、一語ずつ人手によって判断した品詞情報には誤りは殆ど見られなかったが、一律に付与した固有表現情報にエラーが多数出現していた。

例中の“増田順一”は、ゲームソフト“ポケモン”楽曲の作曲家の一人であるが、“ポケモンの関連項目一覧”の記事から取得されており、固有表現情報を記事タイトルから一律

に取得したため、このようなエラーが発生している。対象とした一覧が特定作品などの関連項目の一覧に限定したため、本来であれば、どのように関連しているのかを正しく判定し、その情報を固有表現情報として付与すべきであると考えられる。例えば、見出しとして、“作曲家”などと表記されていれば、その文章を固有表現として利用できたと考えられる。しかしながら、記事によっては、記事中に判断するための情報が不足している場合もあり、今回は一律に固有表現情報の抽出を行った。このため、殆どの関連項目抽出結果について、固有表現情報の付与誤りであると判断された。

このため、手法の見直しと、それに伴う人手による登録された用語の修正を行う必要があると考えられる。

## 7. 終わりに

本研究では、ニコニコ動画のコメントを使用した応用研究のために、コメント内に出現しうる固有表現情報を、ニコニコ大百科における一覧記事から抽出・収録した辞書の構築手法を提案し、実際に構築を行った。その結果、121,359件の固有表現を約8割の精度で抽出・収録した辞書を構築することができた。実用性を考慮すると、収録されている固有表現の数が不足していると考えられるため、今後は一覧記事以外のニコニコ大百科の記事からの固有表現抽出手法の考案が必要である。また、随時新たな固有表現がコメント中に出現していているため、手法の完全な自動化や、コメントデータ自体からの抽出手法も必要であると考えられる。また、構築された辞書を用いて、ニコニコ動画コメントの形態素解析を行うなどの、応用も行い、実用性についても検証を行ってきたい。

謝辞 本研究では、(株)ドワンゴおよび(有)未来検索ブラジルと国立情報学研究所が協力して研究者に提供しているニコニコ動画のデータを活用した。

## 参考文献

- [1] Kazama, J. and Torisawa, K.: exploiting wikipedia as external knowledge for named entity recognition exploiting Wikipedia as External Knowledge for Named Entity Recognition, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 698-707 (2007).
- [2] Siersdorfer, S., Chelaru, S., Nejdil, W. and San Pedro, J.: How Useful Are Your Comments?: Analyzing and Predicting Youtube Comments and Comment Ratings, *Proceedings of the 19th International Conference on World Wide Web*, ACM, pp. 891-900 (2010).
- [3] Wu, B., Zhong, E., Tan, B., Horner, A. and Yang, Q.: Crowdsourced Time-sync Video Tagging Using Temporal and Personalized Topic Modeling, *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, New York, NY, USA, ACM, pp. 721-730 (online), DOI:

- 10.1145/2623330.2623625 (2014).
- [4] 早藤 健, 建石由佳: 2ちゃんねる解析用の形態素解析器の作成, 言語処理学会第16回年次大会 発表論文集, pp. 330-333 (2010).
  - [5] 内山 弘: ネットの日本語-2ちゃんねるとニコニコ動画を中心に-, 鹿児島大学リポジトリ, 地域政策科学研究, Vol. 7, pp. 219-236 (2010).
  - [6] 池田晃人, 小林暁雄, 坂地泰紀: ニコニコ動画のコメントに対する言及内容に基づくアノテーションのための分類, 電子情報通信学会技術研究報告, Vol. 114, No. 211, pp. 47-52 (2014).
  - [7] 中山浩太郎, 原 隆浩, 西尾章治郎: Wikipedia マイニングによる大規模 Web オントロジの実現, 人工知能学会全国大会論文集, Vol. 8, pp. 1-4 (2008).
  - [8] 小林暁雄, 増山繁, 関根聡: Wikipedia と汎用ソーラスを用いた汎用オントロジ構築手法, 電子情報通信学会論文誌 D, Vol. J93-D, No. 12, pp. 2597-2609 (2010).
  - [9] 桜井慎弥, 手島拓也, 石川雅之, 森田武史, 和泉憲明, 山口高平: 日本語 Wikipedia からの汎用オントロジの構築と評価, 人工知能学会全国大会論文集, pp. 1-4 (2008).
  - [10] 大田仁克: Folksonomy を用いた SNS における創造活動の予測, 修士論文, 豊橋技術科学大学 (2015).
  - [11] 伊藤聖修, 鈴木育男, 山本雅人: ニコニコ動画におけるタグ共起ネットワークの特徴抽出, 知識ベースシステム研究会, Vol. 80, pp. 13-18 (2008).
  - [12] 早川卓弥, 土方嘉徳, 西田正吾: 有益な動画コメント抽出に関する基礎検討, 人工知能学会全国大会論文集, Vol. 28, pp. 1-4 (2014).
  - [13] 石坂達也, 山本和秀: 2ちゃんねるを対象とした悪口表現の抽出, 言語処理学会第16回年次大会 発表論文集, pp. 178-181 (2010).
  - [14] 濱野智史: ニコニコ動画はいかなる点で特異なのか: 「擬似同期」「N次創作」「Fluxonomy (フラクソノミー)」, 情報処理学会論文誌, Vol. 53, No. 5, pp. 483-488 (2012).
  - [15] 浜崎雅弘, 武田英明, 西村拓一: 動画共有サイトにおける大規模な協調的創造活動の創発のネットワーク分析-ニコニコ動画における初音ミク動画コミュニティを対象として-, 人工知能学会論文誌, Vol. 25, No. 1, pp. 157-167 (2010).
  - [16] 真大平澤, 祐樹小川, 博彦謙訪, 太田敏澄: ニコニコ動画のログデータに基づくソーシャルノベルティのある動画の発見手法の提案, 情報処理学会論文誌, Vol. 54, No. 1, pp. 214-222 (2013).
  - [17] 佃 洗撰, 中村聡史, 田中克己: 視聴者のコメントに基づく動画検索および推薦システムの提案, 第19回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2011) (2011).