

単語の頻度と意味に基づいた コミックに関するテキスト情報源の特性分析

樋口 亮太[†] 山西 良典^{††} 松下 光範^{††}

[†] 関西大学大学院総合情報学研究科 〒569-1095 大阪府高槻市霊仙寺町 2-1-1

^{††} 関西大学総合情報学部 〒569-1095 大阪府高槻市霊仙寺町 2-1-1

E-mail: †{k896930,ryama,t080164}@kansai-u.ac.jp

あらまし 本稿では、コミックに関する説明文やレビュー文といった同一コンテンツを扱うテキストでも情報源によって記述内容が異なることに着目し、出現単語から各情報源の特性を分析した。コミックは画像と言語によって構成されるマルチモーダルなコンテンツであるため、内容に基づくコンテンツを直接的に処理するアプローチは難しい。間接的にコンテンツの内容を処理するための方法としてコミックに関するテキスト情報の分析があるが、テキストの情報源の違いによる特性についての知見は見当たらない。作品の詳細を伝えるために用意された説明文と作品を読んだ読者の感想が書かれたレビュー文の2種類の情報源の対象とし、テキストに含まれる頻出単語を意味的に分類することで、同一コンテンツに対する異なる情報源での特性の違いを分析した。その結果、説明文には登場キャラクターの家族構成や身体情報、性別といった内容を示す単語が、レビュー文にはイラストや作風といった作品のメタ的な情報を示す単語がそれぞれ特徴的に現れることを確認した。コミックに関連するテキスト情報において、作品の説明文はレビュー文に比べて内容に基づくアプローチの情報源として有用である可能性が示唆された。

キーワード データセットの性質、コミック工学、レビュー文、コミックのキャラクター情報

1 はじめに

1.1 コミックを取り巻く現状

コミックの新刊発行点数は毎年1万点以上にも及び、現在流通しているコミックの総数は膨大である[14]。日々増加を続けているコミックの中からユーザの興味に応じた内容の作品を推薦するためには、コミックの内容に関する情報を把握する必要がある。現状では、この中からユーザが読みたいコミック作品を選択する方法として、電子コミックや書籍販売のウェブサイトといったコミックの検索サービスがある。代表的なコミック検索手法として、技術書や雑誌などの検索と同様に、書誌に関する情報(e.g., タイトル, 作者, 掲載誌)やジャンル(e.g., ラブコメ, アクション・アドベンチャー, ヒューマンサスペンス)といったメタ情報がクエリとして利用されている。例えば、電子書籍・マンガストアの「めっちゃコミック¹」では新着順, 人気キーワードランキングなどの特集ページや, ジャンル別(e.g., 少年漫画, 少女漫画, 青年漫画), カテゴリ別(e.g., ファンタジー, ミステリー, スピンオフ)といったタグ情報から検索する。一方で, コミックは, キャラクターや漫符といった画像情報とセリフやオノマトペといったテキスト情報が相補的かつ協調的に利用されているクロスモーダルなコンテンツであるために, これらのメタ的な情報だけではコミックの内容を表現するには不十分である。

コミックの内容を把握するために, 2種類のアプローチが考えられる。1つ目は, コミック作品の画像を直接的に分析する手法である。上述した通り, コミックというのは異なる形式の情報が混在した複雑なコンテンツであるため, 直接的なアプローチを用いてコミックから内容情報を抽出するためには, コミックのコマの推定[2]やキャラクターの抽出[18], メタデータの付与によるデータセット構築[16]といった様々な手続きが必要となる。現状では, 個々の要素に対する抽出・推定精度は高い値を示す研究も存在するが, これらの要素技術を組み合わせて自動で情報を取得するには再現性の観点で疑問が残る。コミックのアノテーション付きデータセットとして, Manga109[1], [3]などが公開されているが, 最新のコミックに関する内容情報は上述した直接的アプローチを用いることでしか取得できない現状である。2つ目は, コンテンツの内容情報を他のリソースから抽出し, それらの解釈することで間接的に内容情報を把握する手法である。これには, コミック作品に関するレビュー文を分析した研究[5], [20]や, Wikipedia やピクシブ百科事典といった自由参加型ウェブサイトからコミック作品やキャラクターの詳細を説明した文章を分析した研究[4], [9]が挙げられる。他にもコミックに関するテキスト情報として, ウェブ上のQ&Aサイトのコミックカテゴリには質問文が, 上述の「めっちゃコミック」のコミック作品ページにはあらすじ文が存在する。以上のような間接的なアプローチの優れている点は, ウェブ上に膨大な量のデータが存在し, 収集が容易である点が挙げられる。

¹: <https://sp.comics.mecha.cc> (2022年1月6日存在確認)

表 1 説明文の情報ソースとそこから得られた説明文数

自由参加型ウェブ百科事典	説明文数
Wikipedia ²	1,950
ニコニコ大百科 (仮) ³	1,281
ピクシブ百科辞典 ⁴	1,879
アニヲタ Wiki (仮) ⁵	1,140
計	6,250

1.2 解くべき課題

コンテンツ研究を行う上では、同一のコンテンツを対象として複数の情報源が考えられる。例えば、料理情報処理 [8] では料理というコンテンツに関連する情報源として、レシピ情報そのものやレシピに対するレビューなど異なるデータセットが用いられる。また、観光分野 [13] においても同一の宿泊施設についての情報源として、宿泊施設に対する利用者のレビューや宿泊施設の説明文などが考えられる。しかしながら、情報源それぞれの特性を定量的に分析した上で、研究に利用するデータセットについての検討が明示された報告は少ない。

本稿で研究対象として扱うコミックのコンテンツ分析には、数種類 (e.g., レビュー文, 説明文, あらすじ文) の情報源が利用可能と考えられる。これらの異なる情報源のテキストはそれぞれ同一コンテンツを表現している一方で、情報源ごとにその内容は異なると考えられる。例えば、レビュー文ならば作品に対する感想や評価を、あらすじ文や説明文ならば作品の概要についての解説を、それぞれ目的とした記述で構成されている。そのため、それぞれの研究が目指す具体的なアプリケーションに応じて、適切な情報源を利用して研究を行う必要があるが、コミックのコンテンツ分析に用いる情報源の違いによる特性を定量的に分析した研究は見当たらない。コミックに関する情報源の特性を理解することで、コミックの内容情報に関する研究において 1) 情報源の特性の違いに応じた適切な情報源の選択, 2) 共通した性質を持つ情報源の統合によるデータ量の確保, といった利点が考えられる。

本稿では、コミックに関する情報を対象として、同一コンテンツを扱った情報源ごとの特性を分析する。コミックに関する情報源として、コミックの作品の詳細を伝えるために用意された説明文と作品を読んだ読者の感想が書かれたレビュー文の 2 種類を対象とし、頻出単語の意味的な分類を行うことでそれぞれの情報源の特性を明らかにする。

2 分析方法

本稿では、同一コンテンツに対する異なる情報源の特性を分析するために、ウェブ上のテキスト情報から以下の手順で頻出単語の意味的な分類を行う。

- (1) 複数のウェブサイトから収集したコミックコンテンツ

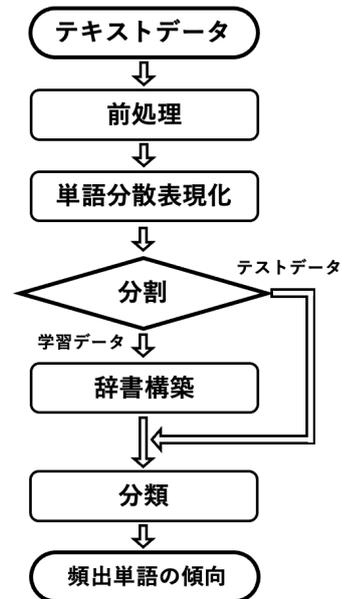


図 1 分析方法のフローチャート

を扱うテキストデータに対して前処理を施し、単語分散表現を用いて出現単語のベクトルを獲得。

- (2) 上記 1 で得られたデータを学習用とテスト用に分割し、学習データを用いてコミックの内容理解のための語彙をまとめたクラス辞書を構築。

- (3) 上記 2 で得られた辞書を用いて、テストデータに含まれる頻出単語を意味的に分類することで、その傾向を分析。

図 1 に、分析方法のフローチャートを示す。

2.1 異なる性質の情報源からのデータセットの構築

コミックに関するテキストデータとして、ウェブ上のキャラクターの説明文と作品のレビュー文それぞれ 6,250 文から成るデータセットを構築した。異なる性質をもつ 2 種類の情報源 (説明文とレビュー文) での記述される内容の性質の違いを分析する。

キャラクターの説明文については、表 1 に示すように、コミックの内容情報が多く含まれている自由参加型ウェブ百科事典を情報ソースとして、コミックのキャラクターについて記述している説明文を 2,067 キャラクター分、計 6,250 件収集した。作品の内容を表現する情報源として、電子書籍・マンガストアの「めっちゃコミック」の各作品のページにはあらすじ文がある。あらすじ文は内容を端的な短文や語句で表現しているため、その作品の特徴となる情報が集約されている点で優れているが、収集が困難である点や情報を集約したために抽象的な表現が含まれている恐れがある。客観的にコミックの内容を表現するためには、一個人の印象ではなく複数人が 1 つの作品に対して自由に書き込んだ情報が集約されたウェブサイトに掲載されている情報が適していると考えた。また、対象を作品の説明文ではなく、キャラクターの説明文にすることで、エピソードレベルで各キャラクターの特徴を抽出可能となり、登場キャラクターに連動して展開される作品の内容をより詳細に捉えられると考えた。

2 : <https://ja.wikipedia.org/wiki/> (2022 年 1 月 6 日存在確認)

3 : <https://dic.nicovideo.jp> (2022 年 1 月 6 日存在確認)

4 : <https://dic.pixiv.net> (2022 年 1 月 6 日存在確認)

5 : <https://w.atwiki.jp/aniwotawiki/> (2022 年 1 月 6 日存在確認)

レビュー文については、「作品データベース⁶」内の漫画カテゴリに分類されている作品の中でも、「最近閲覧数が多い作品トップ1000」から上位200作品を情報ソースとして、そのレビュー本文を計6,250件をスクレイピングで収集した。このウェブサイトは、Amazon.comや楽天市場といったショッピングサイトと異なり、コミック作品の購買が主な目的ではなく、作品の評価・情報集積サイトである。コミックの内容に関する情報を取得する上で、ショッピングサイトからレビュー文を取得するアプローチ[11]も考えられるが、直接的にコミックの内容と関係が小さい情報が混在する恐れがあるため、本稿ではレビューサイトを対象としてテキスト情報を取得した。この時、各作品ごとのレビュー文の数に偏りを出さないために1作品あたりの最大レビュー数の上限を45件とした。

2.2 単語分割とデータクレンジング

テキストに含まれる単語を意味的に分類するための辞書を作成した。本稿で扱う説明文やレビュー文には作品の内容や読者の意見など様々な情報が含まれている。情報源ごとの特性を調査する上で、これらを網羅的に抽出することが望ましい。そこで、品詞の中でも語数が豊富であり、事物や動作、状態といった幅広い意味表現が期待できる名詞に着目した。用意したデータセットの内容を概観したところ、キャラクターの説明文には、“元氣”や“勝利”といったキャラクターの特徴や内容情報に関する単語が、レビュー文には、“王道”や“感動”といった感想に用いられる単語がそれぞれ含まれていた。そこで、情報源ごとの特性を分析するためにテキスト中に含まれる名詞に着目した分析を行うこととした。以上の理由からテキスト中の名詞を対象とするが、一方で、作中に登場するキャラクターの名前や技名といった固有名詞は、内容を把握する上で参考になり得ない情報であるため、このような名詞はあらかじめ分析対象から除外することとする。形態素解析器にはMaCab(バージョン0.996)、辞書にはNeologd(バージョン0.0.7)を用いた。Neologd辞書は、コミックコンテンツを扱うテキストに登場する固有名詞が多数収録されていることを確認した。また、「あれ」や「いつ」といった直接的にコミックコンテンツの意味を示さない語が含まれることを防ぐため、ストップワードを定義した。ストップワードのリストは、Slothlib[7]を参考にし、追加でひらがな一文字とカタカナ一文字、数字や記号も意味の判定が困難なため正規表現を利用して除去している。

情報源ごとに使用されている単語の意味的分類(2.3章で後述)をする上で、低頻度語はノイズになりかねない。そこで、各情報源ごとの特性を分析するために出現頻度が10回未満の低頻度語を除去した。この閾値は経験的に定めた。以上の処理を説明文とレビュー文のデータセットに適用した。その結果、単語の総異なり数は、説明文とレビュー文それぞれ、7,136件と3,092件となった。最後に、2種類の情報源のデータ各6,250件を8:2の割合でそれぞれ分割し、学習データ10,000件、テストデータ2,500件を獲得した。

表2 特徴的な単語が集まったクラスに含まれる単語の一部の例

クラス A	クラス B	クラス C
激戦	黒	アイドル
戦友	白	コンプレックス
初戦	褐色	クラス
苦戦	顔色	ジム
頭脳戦	青春	ポジション
作戦	明白	ピアス
戦国	白衣	バドミントン
対戦	色恋	ジャングル
戦乱	赤血球	ダイビング
戦況	暗黒	リンゴ

2.3 頻出単語の分類

2.2節で前処理を施したデータに対して、単語を意味的に分類する。本稿では、鈴木らが作成したWikiモデル日本語Wikipediaエンティティベクトル[12](以降、Wikipediaベクトル)を用いた。これは日本語版Wikipediaの本文全文を学習データとして構築(ベクトルの次元数が200次元、windowサイズは5単語)されており、一般的な単語や語彙の曖昧性に頑健という特徴を持つ。Wikipediaベクトルを用いて、用意したデータセットをベクトル化した。

得られた単語ベクトルをもとにk-means法を用いて単語を分類した。ここで、最適なクラスター数を求めるためにエルボー法を用いた。エルボー法を用いて2~100クラスまで探索的に調査したところ、63クラスが妥当という結果が得られた。表2に、特徴的な単語が集まったクラスの例を示す。表中、クラスA、クラスBには、それぞれ「戦」や「色」に関する単語が多く分類されている。またクラスCには意味的にではなく、カタカナ語が多く含まれる結果となった。分類の結果、1クラスあたりに含まれる単語数の平均値は118.8ポイント、標準偏差は107.1ポイントとなった。またクラスに含まれる単語の最大値と最小値はそれぞれ495ポイント、8ポイントとなった。ここで得られた単語のクラス集合を、コミックの内容分析に必要な語彙をまとめたクラス辞書として扱う。

コミックの内容分析では、図2に示すように、クラス辞書に含まれる63クラスに該当する単語が用いられる傾向を考察する。63クラスに含まれる単語群を手がかりとして、テストデータ2,500件に出現する単語について各データでの出現頻度を算出する。例えば、あるテストデータに“リンゴ”という単語が含まれていた場合、表2よりクラスCに同じく“リンゴ”とあるので、このデータはクラスCの要素を含むとして、1を立てる。2,500件のテストデータそれぞれについて63要素のバイナリ列が得られる。

3 結果と考察

情報源ごとに多くのデータが該当したクラスを分析したうえで、2種類の情報源の性質を比較するために各クラスの出現が占める割合を算出して考察した。以下では、各クラスごとの情報源における割合の差の絶対値を相対差として定義する。

6: <https://sakuhindb.com/> (2022年1月6日存在確認)



図2 コミックの内容理解のための語彙をまとめたクラス辞書を用いた頻出単語の意味的分類

表3 説明文の頻出単語とレビュー文と比較した割合の差

クラスに含まれる単語群	割合の差
身, 身長, 上半身, 変身, 身分, 身近	74.2
条件, 無条件, 信条, 戸, 票	69.0
親, 兄, 妹, 姉, 孫	63.7

表4 レビュー文の頻出単語と説明文と比較した割合の差

クラスに含まれる単語群	割合の差
漫画, 映画, イラスト, 画風, 原稿	35.5
作品, 作家, 作画, 傑作, 駄作	19.8
節, 巻, 世界, 続編, 世代	3.1

3.1 説明文の特性

説明文には、キャラクターの特徴や作品の内容を説明するような単語が多く出現した。表3に、レビュー文と比較して説明文の割合が多く含まれていたクラスを示す。

最も大きな相対差があったクラスは、身長や身近といった「身」という語が含まれる単語群が多く分類されていて、説明文の相対差は74%であった。「鬼滅の刃」に登場する「竈門炭治郎」のWikipediaの項目に「身長は165cm」と記述があるようにキャラクターの説明文には、身長に関する情報が高頻度で含まれているためにこのような結果になったと考えられる。また、「身近」という語はキャラクターの周りで起きたエピソードを説明するために用いられる場合が多く、合わせてこのクラスが高い割合で説明文に含まれる情報を内包している要因になったと考えた。

次に大きな相対差があったクラスは、条件や信条といった単語を含むクラスである。このクラスは、含まれる単語数が8単語と全クラスの中最も単語の数が少なかったにもかかわらず、説明文の割合は69%と高い割合を示した。このクラスの単語が含まれている説明文は、「東京喰種」に登場する「金木研」のpixib百科事典の項目には、「交換条件により仲間の助命と脱走を求め受け入れられるものの自身はアオギリ内に戻りヤモリの趣味である拷問を受けることとなる。」というように作品の内容に関する情報を記述している傾向があった。レビュー文には、ネタバレ[10]を避けるために作品の内容そのものには直接触れない場合もあるため、エピソードを説明している部分に登場する単語は、説明文のほうに多く登場する傾向にある。また、条件という単語はキャラクターの能力に関する記述の中にも頻出する傾向にあった。「鋼の錬金術師」に登場する「ロイ・マ

スタング」のアニヲタ Wiki (仮)の項目には、「雨が降りしきる屋外などでは、条件が厳しくなりすぎて無力化してしまう欠点を持つ。」という記述がある。

3番目に多くのデータに該当したクラスは、家族構成を意味する単語 (e.g., 親, 兄, 孫) が多く含まれていた。これらの単語は、キャラクタ間の関係性[17]を説明する際に多く用いられており、「ONE PIECE」に登場する「サボ」のピクシブ百科事典の項目には、「エースやルフィと共に幼少期を過ごした、ルフィのもう一人の義兄。」という記述がある。

3.2 レビュー文の特性

レビュー文の特性としてコミックのイラストや作風、作者といった作品のメタ的な情報を示す単語群が含まれていた。表3に、説明文と比較してレビュー文の割合が多く含まれていたクラスを示す。

実際のデータには、「画風、熱量といいスポーツ漫画に限らないと言える後世に与えた影響が大きい作風。」というように作品の内容に関する情報よりも、レビュー文の著者が作品を読んでどのような感情を持ったかについて解釈に関する情報が多く含まれていた。このことから、作品のカテゴリを表現するジャンル情報に関する研究[19]やトピック分類[6]などに利用できる可能性が示唆された。一方で、最も大きな相対差があったクラスでも35.5%と、説明文と比べて相対差が大きく異なる結果とはならなかった。これはレビュー文のデータの中に、「面白かった。また読みたい。」のような非常に短いデータも含まれてしまっていることが原因の1つだと考える。

説明文に対して大きく割合の差があるクラスは存在せず、最も大きな差が見られたクラスでもその値は約40%未満にとどまった。情報源ごとに該当したクラスの合計値を算出すると、説明文は35,694ポイントに対して、レビュー文は20,429ポイントと説明文データの方がレビュー文と比べて約1.75倍のクラスが該当していた。この原因として考えられるのは、形容詞を対象外としてしまった点が挙げられる。レビュー文には、「ハラハラドキドキの展開がアツい」や「表情が怖かった」のような形容詞を用いた表現が散見された。本稿では、一般名詞のみを対象としたため、このような形容詞を用いた作品に対する感想の情報は除外されてしまっていることが要因と考えられる。分析対象とする品詞を変えた分析と考察については、今後の展望とする。

表5 最も多くのデータが該当したクラスに含まれる単語群

ヘアスタイル, チェック, ブラモ, リハビリ, キャラ, ヒップ, ウエスト, アニメ, ギャル, ダイエット, バイト, ヤクザ, ケーキ, テレビ, バイク, アップ, マスコット, タイヤ, キャラクター, ピアノ, ピアニスト, タイプ, インターハイ, キャプテン, バカ, モテモテ, バスケット

3.3 全体の考察と展望

63クラスの中で最も多くのデータが該当したクラスは、表5に示すようなカタカナ語を含んだクラスで全体の73%のデータが該当した。このようなカタカナ語が集中したクラスは他にも2件存在しており、これらのクラスは意味的な分類ができていない。これらの分類に関しては異なるコーパスを用いることで改善が期待できる。

本稿では、Wikipediaベクトルを用いて単語を分散表現化した。それに対して、松野ら[15]は、ブログやTwitter等のSNSデータ、日本語Wikipedia、自動収集したウェブページなどを中心とした日本語コーパスを構築した。このコーパスは、本稿で対象としたキャラクタの説明文と作品のレビュー文に多く含まれているネットスラングや砕けた表現などの新語・未知語を多数収録しているため、表5のようなカタカナ語の意味を解釈することが期待できる。

4 おわりに

本稿では、コミックコンテンツを扱うテキスト情報源の中でも説明文とレビュー文を対象とし、単語の出現頻度から情報源の特性を分析した。その結果、説明文はキャラクタの特徴や作品の内容を説明するような単語が頻出し、コミック作品の内容に関する分析に適した情報源であることが示唆された。また、レビュー文はコミックのイラストや作風、作者といった作品のメタ的な情報を示す単語群が含まれており、ジャンル推定などの課題において有用性が期待される情報源であることが示唆された。

今後の検討課題としては、分析対象として着目する単語の検討が必要であると考えられる。本稿の分析では、一般名詞のみを分析対象としたため、レビュー文に含まれる“アツイ”や“怖い”といった作品に対する評価を示す形容詞を用いた感想は除外されてしまった。今後は、形容詞や動詞といった品詞も対象とすることで、コミック作品に関するテキストの情報源の特性を詳細に分析していく。

謝辞

この研究は2021年度国立情報学研究所公募型共同研究(21S0501)の助成を受けています。

文献

[1] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T. and Aizawa, K.: Sketch-based manga retrieval using manga109 dataset, *Multimedia Tools and Ap-*

plications, Vol. 76, No. 20, pp. 21811–21838 (2017).
 [2] Nguyen Nhu, V., Rigaud, C. and Burie, J.-C.: What do We Expect from Comic Panel Extraction?, Vol. 1, pp. 44–49 (2019).
 [3] Ogawa, T., Otsubo, A., Narita, R., Matsui, Y., Yamasaki, T. and Aizawa, K.: Object detection for comics using manga109 annotations, *arXiv preprint arXiv:1803.08670* (2018).
 [4] Park, B., Kanae, I. and Mitsunori, M.: Classifying Personalities of Comic Characters Based on Egograms, *International Symposium on Affective Science and Engineering ISASE2018*, Japan Society of Kansei Engineering, pp. 1–6 (2018).
 [5] Ueno, A., Kamoda, Y. and Takubo, T.: A spoiler detection method for japanese-written reviews of stories, *International Journal of Innovative Computing Information and Control*, Vol. 15, No. 1, pp. 189–198 (2019).
 [6] Xu, A., Qi, T. and Dong, X.: Analysis of the Douban online review of the MCU: based on LDA topic model, Vol. 1437, No. 1, p. 012102 (2020).
 [7] 大島裕明, 中村聡史, 田中克己: Slothlib: web サーチ研究のためのプログラミングライブラリ, *DBSJ Letters*, Vol. 6, No. 1, pp. 113–116 (2007).
 [8] 菊地悠樹, 熊野雅仁, 木村昌弘ほか: 料理レシピ共有サイトにおける食材のアクティブ共起パターンの抽出, 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 11, No. 2, pp. 30–40 (2018).
 [9] 小林達哉, 松下光範: 性格要素と外見要素の加減算による類似キャラクタの検索, 第16回 ARG WEB インテリジェンスとインタラクション研究会予稿集, pp. 106–111 (2020).
 [10] 佐藤剣太, 牧良樹, 中村聡史: 未読および既読シーンの提示が読者のコミック閲覧意欲に与える影響, 情報処理学会研究報告, Vol. 2018-EC-47, No. 3, pp. 1–8 (2018).
 [11] 白石絵里奈, 田村亮介, 浅沼爽汰, 白井聡一, 藤田和成, 町田翔, 延澤志保: 概要と読者レビューに基づく漫画特徴抽出, 情報処理学会第81回全国大会講演論文集, No. 1, pp. 443–444 (2019).
 [12] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎: Wikipedia記事に対する拡張固有表現ラベルの多重付与, 言語処理学会第22回年次大会発表論文集, pp. 797–800 (2016).
 [13] 野守耕爾, 神津友武: 観光に関するユーザーレビューデータを用いた観光客の話題分析と地域観光振興への活用の検討, *サービソロジー論文誌*, Vol. 2, No. 2, pp. 1–12 (2019).
 [14] 堀川喜一: 出版月報2月号, 公共社団法人全国出版協会出版科学研究所 (2020).
 [15] 松野省吾, 水木栄, 榎剛史: 日本語大規模 SNS+Web コーパスによる単語分散表現のモデル構築, 第33回人工知能学会全国大会論文集, 4Rin113 (2019).
 [16] 三原鉄也: マンガの知的利用のためのメタデータ, *日本画像学会誌*, Vol. 58, No. 5, pp. 538–544 (2019).
 [17] 森理緒奈, 山西良典, 松下光範: 発話の役割ベクトルによる登場人物間の有向関係表現の有用性検証, 電子情報通信学会第5回コミック工学研究会, pp. 55–61 (2021).
 [18] 柳澤秀彰, 山下拓朗, 渡辺裕: 主要キャラクタの抽出を目的とした漫画キャラクタ画像のクラスターリング, *映像情報メディア学会誌*, Vol. 73, No. 1, pp. 199–204 (2019).
 [19] 山下諒, 松下光範: コミックの探索的検索支援にむけたジャンル情報の利用に関する検討, 第29回人工知能学会全国大会論文集, 1H23in (2015).
 [20] 山下諒, 朴炳宣, 松下光範: コミックの内容情報に基づいた探索的な情報アクセスの支援, *人工知能学会論文誌*, Vol. 32, No. 1, pp. WII–D-1 (2017).