

執筆者属性に基づく日記文章のトピック特徴の抽出

松田 雛乃・村井 源（公立はこだて未来大学大学院 システム情報科学研究科）

概要：本研究では日記の重要性に着目し、日記文章の自動生成を目指す上で有用となるトピック特徴の分析を行った。内容を確認して日記文章と判断できるブログ記事を抽出し、執筆者属性の影響を想定して老年男性・老年女性・若年男性・若年女性の4カテゴリにおいて総計80人分のブログ記事を収集した。収集した文章をトピック単位に分割した上でトピックの内容を表わすタグを各トピックに付与し、タグの出現頻度を用いてカイ二乗検定による分析を行った。結果から、男女や年齢層での差異や、老年男性・老年女性・若年男性・若年女性の執筆者属性ごとのトピック特徴が判明した。

キーワード：日記, トピック抽出, カイ二乗検定

Extraction of topic features of diary text based on author attributes

Hinano Matsuda / Hajime Murai (Graduate School of System Information Sciences, Future University Hakodate)

Abstract: In this study, we focused on the importance of diaries and analyzed the topic features that are useful for automatic generation of diary texts. The blogs that were similar to the diary texts were substituted as the diary texts by actually reading through and checking the contents, and a total of 80 blog posts were collected in the four categories of old men, old women, young men, and young women, assuming the influence of author attributes. The collected texts were divided into topic units, and tags representing the contents of the topics were assigned to each topic, and the frequency of occurrence of the tags was used for analysis by the chi-square test. The results revealed differences in gender and age groups, as well as topic characteristics for each author attribute: old men, old women, young men, and young women.

Keywords: Diary, Topic Extraction, Chi-squared Test

1. まえがき

日記は、史料としての価値や文学作品としてのエンタテインメント性[1]や、行動や感情を把握する記録媒体[2][3]としての機能性など様々な要素を持ち合わせており、読み物・書き物の両面において古来より人々に活用されている。

多くの日記は決まった形式がなく、書き手がその日行ったこと・考えたことを自由に書き記す口語調に近い文章である。そのため、年齢や性別などの書き手の属性の違いがトピックに影響を及ぼす可能性が考えられる。よって、それぞれの属性における日記のトピックの特徴を分析することで、ブログの著者属性の推定や人の興味関心を考慮したマーケティングなどにも応用できると考えられる。また、分析結果を日記文章の自動生成や日記執筆のサポートツールなどに用いることで、日記をより手軽なものとし、多忙な中でも日記を書きたいと望む人々の支援に繋がると考えられる。

筆者らはこれまで、実際の日記文章の特徴に基づいた日記文章のプロット自動生成を目標とし、書き手の属性による表現や内容の差に着目した研究を行ってきた。対象データとして、読者から人気の高いブログの記事を収集して用いている。文章表現に関しては、機能語に着目した文体特徴の分析を行い、属性の異なる書き手の日記文章における文体特徴の差を抽出した[4]。また、

日記文章の論理的な構成を明らかにするため、文の機能を記号化し、その時系列的な特徴を n-gram 分析により明らかにした[5]。本研究では、日記文章の内容に着目し、文章をトピック単位で区切った上で書き手の属性ごとにトピックの特徴の分析を行い、書き手の属性に対応した日記文章のプロット自動生成に必要な基礎データを収集することを目的とする。

2. 分析対象

本研究では、書き手の年齢や性別などで特徴が変化する可能性を考慮した上で日記文章を分析する。実際の日記文章の収集の難しさや個人情報保護の観点から、データソースとしてブログからの収集を行った。

ブログに関する自然言語処理の先行研究として南野らの blog の自動収集と監視[6]が挙げられる。ブログのマイニングや評判分析等のマーケティングでの利用は行われてきているが、ブログ自体が何をどのように語っているのかは分析の対象となっていない。

ブログはブログランキングサイトの「にほんブログ村」[7]から、カテゴリ機能を用いて書き手の情報を参照し、男・女の性別と若年・老年という年齢層の二つの属性を組み合わせて4カテゴリに分類した上で、ある二つの条件を満たすものを上位20サイトまで選定を行った。なお、若年の年齢層は10~20代、老年の年齢層は50代後半以

上である。一つ目の条件は、そのブログの内容が、書き手がその日に行ったことあるいは考えたことを題材にしたもの、つまり日記に近いブログであると判断できることである。この条件を設けることで、ブログ記事を日記に類似した文章として扱った。二つ目の条件は、ブログ内で可視のエントリ数が 20 個以上存在することである。この条件は、特に年齢層の若い書き手のブログにおいて、すぐに更新されなくなる現象が発生していることを鑑み、収集・分析の対象であるブログの書き手一人一人のデータ数を十分に確保するために設けている。そのため、条件を満たさないブログはランキング上位の場合でも除外した。

表 1 は、対象データの各カテゴリにおいて選定した書き手の人数と、データサイズとブログエントリの数、つまりファイル数の内訳と合計である。

表 1 分析対象データ詳細

カテゴリ	人数	データサイズ	ファイル数
老年男性	20	52.1MB	43120
老年女性	20	64.6MB	27506
若年男性	20	10.7MB	4534
若年女性	20	9.5MB	4642
合計	80	136.9MB	79802

収集した表 1 の日記文章から、本研究において分析対象となるファイルのランダムサンプリングを行った。各カテゴリ 20 人、合計 80 人の書き手から 5 ファイルをランダムに選定し、総計 400 ファイルを用いて文章構造の分析[5]と本研究のトピック特徴の分析を行った。

3. 方法

ランダムに選定し、文章構造の分析[5]を行った 400 ファイルを用いて、トピック種別のタグ付けを行った。文章構造の分析[5]のために行った手法と合わせ、本研究におけるトピック種別のタグ付けの手順を説明する。

1. 一文ごとの区切り

まず、収集したブログの本文を一文ごとに区切った。基準は以下の通りである。

- ・「。」「？」「！」など、句点の機能を持つ記号で区切る
- ・句点の機能を持つと判断できる場合に限り顔文字や絵文字で区切る
- ・括弧で区切る
- ・句点に成り得るものがない場合、文として不自然な途切れがない限り改行で区切る

なお、「？」「！」や括弧については、一文の効果・セリフ・補足などに使われていると判断できる場合は区切らないこととした。

2. 文の機能のタグ付け

区切った一文ごとに対し、その一文が文章内で果たす機能のタグを付与した。機能の一覧を以下の表 2 に示す。

表 2 文の機能のタグの定義表

タグ名	定義
トピック提示	トピックを提示する文
トピック掘下	トピックを説明・補足などで掘り下げる文
意見感想	トピックについて、書き手が考え・気持ちを述べる文
名乗り／挨拶	書き手の名乗りまたは挨拶の機能を持つ文
まとめ	複数トピックを総括するトピックの始点の文に付与
和歌	短歌・俳句・川柳など、特定の形式を持つ詩
その他	上のどれにも当てはまらない文

「トピック提示」は、まず、冒頭の一文が意味不明でなく、かつトピックの内容として理解できる場合に付与した。文章が続くなかで新たな話題に切り替わった場合には新たな話題を示す一文に付与し、また、トピックについて掘り下げるなかで出現した新たな要素に対しても掘り下げが行われた場合、その新たな要素を最初に含んだ一文に付与した。また、新たな要素もなく、先に出現した「トピック提示」の文と全く同一の内容の文に対しても「トピック提示」の付与を行った。また、トピックとして意味は読み取れるものの、提示されるトピックの断片であると判断できる文にも「トピック提示」の付与を行った。

「トピック掘下」と「意見感想」は、両者とも「トピック提示」において示されたトピックについて述べる文へのタグであるが、「トピック掘下」は書き手の主観を含まない事実である文に付与し、「意見感想」は書き手の考え・感情に基づく文に付与した。「トピック掘下」と「意見感想」の判断が曖昧である文が出現した場合、本研究の対象データがネットから収集した文章であり、読み取れる書き手の意図は想像上のものであることを加味し、以下の判断基準を設けた。読み手側からの事実検証が不可能である文においては、荒唐無稽な内容ではなく、過剰な表現や悪意ある装飾がなされていない文には「トピック掘下」を付与し、それ以外の文に「意見感想」を付与した。

「まとめ」については、あるトピックがそれ以前に登場した複数のトピックの内容を総括し、書き手が考えや感想を述べていると判断できる場合に、「トピック提示」と同様にトピックの始点として付与した。

「名乗り／挨拶」については、書き手の名乗り・自己紹介のほか、「おはよう」「こんにちは」「こんばんは」などの出会いの挨拶やそれに類似

した言葉や、「さよなら」「おやすみ」「いただきます」などの別れの挨拶を含む文に対して付与した。また、ブログ特有の要素である「ようこそ」「おかえり」などの意味合いの挨拶を含む文も「名乗り／挨拶」タグを付与した。

3. トピック ID の付与

次に、トピックの提示順と区切りを示すトピック ID の付与を行った。基本的に「トピック提示」「まとめ」というタグを付与した文を基準に ID を付与した。「トピック提示」「まとめ」の出現順に、その「トピック提示」「まとめ」に関連した意見や感想、説明であると判断できる各文に1から番号を付与し、新たなトピックが提示された場合には2番,3番...と出現順に ID を付与した。

「トピック提示」タグを基準にしない例は、ある一つのトピックが「名乗り／挨拶」「その他」「和歌」で始まっている場合と、一つのトピック内で先に出現した「トピック提示」の文と全く同一の内容と判断して「トピック提示」を付与した文が含まれる場合と、トピックの断片であるとして「トピック提示」の付与を行った文が含まれる場合である。「名乗り／挨拶」「その他」は、直後で「トピック提示」を付与された文と同じ ID を付与し、「和歌」は文中の詩に関連性のあるトピックを含む「トピック提示」と同じ ID を付与した。同一内容あるいは断片の文に対しての「トピック提示」については同一内容・断片の大元の「トピック提示」と同じ ID を付与した。

分析対象から除外する文に対してのトピック ID には 0 を付与した。分析から除外する基準は以下の通りである。

- ・ ブログについての応援や読者登録などを読者に願う文
- ・ 記号のみの羅列
- ・ 意味が読み取れない文

なお、ブログを書くことそのものについての意見やアドバイスなどが主題となっている場合、トピック ID の付与対象としては除外しないこととした。

ここまでの 1. 2. 3. の手順は文章構造の分析[5]での手法と同一である。この手順で作業を行ったデータの一部を表 3 に示す。

続いて、本研究で行ったトピック種別のタグ付けの手順を説明する。1,2,3...と続くトピック ID が示すトピック一つ一つに対し、そのトピックにおいて書き手が何を題材としているのかを表わすタグの付与を行った。なお、トピックの内容を表わすタグを付与するにあたっては、実際の分析対象の文章を通読することでタグの収集、分類と命名を行った方が収集データとマッチし、分析の効果も高いと判断したため、日記のテーマについ

ての何らかの先行研究や書籍等は参考にしていない。表 4 にトピック種別のタグとその定義を一覧として示す。トピック種別のタグを付与した 400 ファイルを用い、各種分析を行った。

まず、老年男性・老年女性・若年男性・若年女性のそれぞれのカテゴリにおける、トピック種別のタグの出現頻度をカウントした。その結果、4 カテゴリのうち 3 カテゴリで出現頻度がゼロであり、出現した 1 カテゴリにおいても出現頻度が 10 未満と低いものであったタグを除外した上で、各タグ 29 個と年齢性別 4 種でカイ二乗検定を行った。除外したタグは訃報・もしも話・運動する・スピリチュアル・不明の 5 個である。

表 3 若年男性の日記文章に記号化と ID 付与を行った結果(一部抜粋)

本文	機能	ID
勉強勉強日記	トピック提示	1
本当はね、駿台高2全国模試の結果を乗せようと思ったんですが、何か見れないので春休みの日常の話(新古今和歌集の省略)	トピック提示	1
たぶんね、web返却+紙返却ではなくて、紙返却だけを選んでしまったんだと思います。	意見感想	1
痛恨のミス。	意見感想	1
さて、本題の今日やったこと。	トピック提示	2
今日は掃除した。	トピック掘下	2
以上。	その他	2
一応毎日、最低限これはやろうと思ってることはあって、[1]数学の問題集2ページ[2]英語の長文1つ[3]入門英文問題精講(竹岡広信)[4]黒本国語1問[5]単語王2ユニット[6]英語のなんかを聞くですね。	トピック提示	3
明日からは頑張りますたぶん	意見感想	3
最後に一押しお願いします()にほんブログ村大学受験ランキング	その他	0

4. 結果

カイ二乗検定の結果、トピックタグと各カテゴリのベクトルにおいて有意差が認められた($p < 0.05$)。以降の表 5 にその結果を示す。表 5 中で▲と▲▲はそれぞれ、5%有意水準、1%有意水準で多い個所を示す。同様に▽と▽▽はそれぞれ5%有意水準、1%有意水準で少ない個所を示す。また、各タグと各カテゴリに対応する数値は、そのタグが何回出現したか、つまりそのタグが表わすトピックが各カテゴリで何回出現したかという頻度であり、本研究においてカイ二乗検定に用いた値である。

表4 トピック種別のタグの定義表

タグ	定義	タグ	定義
日常	以下のタグにあてはまらないが、日常を書き綴っているトピック	季節天候	季節天候, 災害についてのトピック
押し語り	書籍や映像作品, 芸能人やキャラクターに対してのトピック	恋愛	恋愛についてのトピック
ブログ	ブログを書いていること自体がテーマになっているトピック	お金経済	貯金や経済的不安などのトピック
病気	健康への不安や, 病気・怪我などについてのトピック	習い事	何らかの習い事・所属クラブについてのトピック
勉強	学校や受験, 資格勉強についてのトピック	ニュース	政治・時事問題などのニュースについてのトピック
料理食事	料理をしている, 食事をしているトピック	哲学悩み	人生や何らかの話題についての思索のトピック
家族	家族親戚についてのトピック	仕事	仕事やアルバイトについてのトピック
訃報	訃報についてのトピック	買い物	買い物についてのトピック
友人	友人知人についてのトピック	掃除	自宅の掃除についてのトピック
学校	書き手の通う学校について, あるいは学校での出来事のトピック	運動観戦	スポーツを実地で観戦, あるいはテレビで観戦するトピック
芸能	芸能人・芸能界についてのトピック	運動する	書き手自身がスポーツをしているトピック
庭仕事	畑や花などの育成についてのトピック	スピリチュアル	超自然的で不思議な話, あるいは怪談・オカルトなどの霊的な話などを書き手の実体験として書いたトピック
ホビー	趣味嗜好を持った玩具やゲームなどについてのトピック	旅行	遠出や旅行などで観光しているトピック
もしも話	アイデアや想像の話を披露するトピック	ツール	趣味嗜好の要素を含みつつ, 実生活や仕事などで役立つ実用的なもの(文房具・デバイスなど)についてのトピック
イベント	参加したイベントについてのトピック	うんちく	興味を持った事象についての知識や思索を披露するトピック
就活	就活についてのトピック	トラブル	他人との諍いや機器の故障など, なんらかのトラブルについてのトピック
懐古	昔を懐かしむ, あるいは書き手の武勇伝的なトピック	不明	何をトピックとしているか読み取れないもの

5. 考察

1. 男女差

老年男性と若年男性の男性 2 カテゴリと老年女性と若年女性の女性 2 カテゴリとに分割し, 日記文章のトピックにおける男女差について考察する. 表 5 から, 男性 2 カテゴリは「うんちく」のトピックの出現頻度が高い一方で「友人」のトピックが少なく, 女性 2 カテゴリは「ニュース」と「うんちく」のトピックが低いことが読み取れる. また, 老年男性が「懐古」「ニュース」などの出現頻度が高く, 若年男性では「ブログ」「ホ

ビー」などの出現頻度が高いことや, 老年女性では「家族」「友人」, 若年女性では学校での出来事・人間関係のトピックである「学校」が多く出現していることから, 日記のトピックには男女差の存在が推測される.

老年男性の「懐古」は思い出を振り返りつつも武勇伝的に自分を語るトピック, 「ニュース」は政治や時事問題などを取り上げて意見・批判を述べるトピックであり, どちらも自身の興味や思考・意見を中心としたトピックである. また, 若年男性の「ブログ」はブログを書くことやブログの書き手としての自分自身を主題としたトピッ

ク、「ホビー」はゲームや玩具など趣味嗜好を含んだ娯楽のトピックである。こちらも自分自身の興味や思考意見を中心としたトピックといえる。

表5 著者属性とトピックのカイ二乗検定

	老年男性	老年女性	若年男性	若年女性
日常	▽ 9	23	11	▲ 26
押し語り	22	▽▽ 4	17	▲▲ 43
ブログ	9	▽▽ 3	▲▲ 29	14
病気	22	▲▲ 52	▽▽ 7	41
勉強	▽▽ 0	▽▽ 0	▲▲ 65	▲▲ 48
料理食事	28	▲▲ 60	▽▽ 14	36
家族	▽▽ 9	▲▲ 96	▽▽ 0	▽ 25
友人	▽▽ 0	▲▲ 25	▽▽ 2	10
学校	▽▽ 0	▽▽ 0	25	▲▲ 59
芸能	▽▽ 2	▽▽ 0	3	▲▲ 18
庭仕事	▲▲ 43	▲ 29	▽▽ 0	▽▽ 9
ホビー	▽▽ 2	▽▽ 1	▲▲ 61	▽▽ 1
イベント	3	4	2	0
就活	▽▽ 0	8	▲▲ 15	▽ 1
懐古	▲▲ 26	11	▽▽ 0	7
季節天候	▲▲ 41	▲▲ 34	▽▽ 2	▽▽ 8
恋愛	▽ 0	▽ 0	4	▲▲ 9
お金経済	4	▲▲ 20	▽▽ 0	▽▽ 1
習い事	9	7	4	13
ニュース	▲▲ 76	▽▽ 22	30	▽▽ 12
哲学悩み	▲▲ 22	9	7	▽▽ 12
仕事	6	13	4	▲ 16
買い物	2	5	▲▲ 14	4
掃除	6	4	4	▽ 0
運動観戦	▲▲ 15	2	3	2
旅行	20	▽▽ 7	▲▲ 42	28
ツール	▽ 11	▽▽ 8	25	▲▲ 38
うんちく	▲▲ 51	▽▽ 13	▲ 39	▽▽ 16
トラブル	12	▲▲ 23	4	▽▽ 0

一方の女性2カテゴリでは、老年女性の「家族」「友人」や若年女性の「学校」はどちらも自分とその人間関係を題材としたトピックである。よって、男性は自分を中心に興味がある物事について語り、女性は自分と人との関わりを含んだ物事について語るという傾向があると考えられる。

2. 年齢差

続いて、老年男性と老年女性の老年層2カテゴリと若年男性と若年女性の若年層2カテゴリから、日記文章のトピックにおける年齢差について考察する。

表5から、老年層2カテゴリでは「庭仕事」「季節天候」の出現頻度が高い一方で、「学校」「勉強」が低い。若年層2カテゴリでは「勉強」の出現頻度が高い一方、「庭仕事」「季節天候」「お金経済」が低いことなどから、年代によるライフスタイルの差異がトピックに影響を及ぼしていると考えられる。ある程度の時間と手間が必要な「庭仕事」や、日々の季節と天候を感じる「季節天候」が老年層2カテゴリで出現頻度が高い点は、仕事や子育てなどを終えて日々の時間にゆとりがあるためと考えられる。学校に通う、友人と遊ぶなどして定期的に家から離れることの多い若年層とは対照である。

若年層2カテゴリで「庭仕事」「季節天候」の出現頻度が低い理由は、前述の通り老年層2カテゴリと比べて時間にゆとりがないためと考えられ、「お金経済」の出現頻度が低い理由として考えられるのは若年層2カテゴリの書き手はほぼ全員が保護者に養育される立場にあり、自身の生活を懸けて真剣な経済をする必要が少ないためと考えられる。

3. 老年男性カテゴリの特徴

老年男性カテゴリにおいては「懐古」「ニュース」などのほか、「哲学悩み」「運動観戦」の出現頻度が高い。「哲学悩み」はその日気付いた物事を取り上げて思索を深めるトピックで、「運動観戦」はスポーツ観戦を題材としたトピックである。なお、「運動観戦」には現地での観戦のほかテレビやラジオなどを通じて知った試合結果を題材にして語ることも含めているが、全カテゴリを通して現地での観戦は1例のみであり、他はテレビからの情報か試合結果のみを述べて情報源は不詳であった。老年男性カテゴリでは、主にテレビを情報源として興味を刺激されて政治や企業や時事問題などに意見を述べるが多かった。また、自身の記憶を振り返る「懐古」や、テレビからの刺激や自身の記憶などから題材を取り上げて知識を披露する「うんちく」が多かった。よって、老年男性はテレビを情報源に発展するトピックと、自身の記憶の振り返りから発展するトピックが特徴であると考えられる。

4. 老年女性カテゴリの特徴

老年女性カテゴリは「家族」「友達」などのほか、「病気」「料理食事」「お金経済」「トラブル」

の出現頻度が高い。「家族」の出現頻度が他のカテゴリと比べて高い理由としては、自身が夫や息子・娘、孫と過ごす時間が長く、日記の題材となりやすいためと考えられる。また、「料理食事」が高い理由も、書き手自身が食卓を預かる立場である割合が高いためと考えられる。「病気」については、書き手が高齢であるため健康上の不安や問題が発生しているために出現しやすいと考えられ、「お金経済」も食費等の生活費を気にかけるべき立場である割合が高いためと考えられる。「トラブル」は家電の故障や隣人との諍いなどの非常事態を題材としたトピックであり、これも家を預かる立場である割合が高く、非常事態に直面する可能性が高いためではないかと考えられる。老年女性は自身の健康を気にかけるトピックや、自身の人間関係を含んだ日常を題材にしたトピックが特徴であると考えられる。

5. 若年男性カテゴリの特徴

若年男性カテゴリでは「勉強」「ホビー」などのほか、「就活」「旅行」の出現頻度が高い。「旅行」は修学旅行などの観光地等に外出する学校行事を含んでおり、「勉強」は国語数学等の科目の勉強や受験の情報を含んでいる。「就活」は大学生の書き手が企業分析などの就職活動を題材としている。そういった学校や自身の将来に関係した出来事を語るトピックのほか、「ホビー」など趣味や知識、自分自身について語るトピックが多く出現した。よって、若年男性は学校や自身の将来を題材にしたトピックや書き手の趣味・知識を題材にしたトピックが特徴であると考えられる。

6. 若年女性カテゴリの特徴

若年女性は「学校」のほか、「推し語り」「芸能」「恋愛」「仕事」「ツール」の出現頻度が高い。「推し語り」は本やドラマ・アニメなどに登場する俳優やキャラクター、あるいは作品自体について語るあるいは紹介するトピックであり、「芸能」は作品を介さない視点から芸能人や芸能界について語るトピックである。また、「恋愛」は字で示す通り書き手自身の恋を題材にしたトピック、「仕事」は正社員以外にアルバイトも含めての仕事やアルバイトを題材にしたトピック、「ツール」は趣味嗜好を含みつつも実用的な品について語るあるいは紹介するトピックである。なお、「仕事」について若年女性カテゴリにおいては全ての実例がアルバイトの話題であった。

よって、若年女性は自身の趣味嗜好を他者に紹介するトピックと、自身のアルバイト経験を語るトピック、自身の人間関係・恋愛関係を語るトピックが特徴であると考えられる。

6. 今後の展望

本研究は男性・女性と老年・若年の属性を組み合わせた4カテゴリの書き手の日記文章を収集し、ランダムに選出された400ファイルに対して一文の区切りやトピックIDの付与、またトピック種別のタグ付け等を行い、付与したトピックタグの出現頻度を用いてカイ二乗検定を行った。その結果、性別と年代での共通点や差異、各カテゴリにおけるトピック特徴が判明した。

課題として、先行研究[5]と本研究で行った区切りや各種タグの付与、番号付与等は単独の分析者で行っているため、定義表をもとに複数人での一致度検証を行い、客観性の担保を行う必要がある。トピック種別のタグについても実際の文章の内容に適したタグが付与されているかを検証し、客観性を担保する必要があるといえる。

今後の展望としては、トピックからトピックへの接続についての傾向の分析を行い、本研究の成果を先行研究[4,5]での文体分析の結果と合わせることで日記文章のプロット自動生成を実現することなどが挙げられる。

参考文献

- [1] Webster, J: Daddy-Long-Legs, 1915. (岩本正恵訳: あしながおじさん, 岩波書店, 1970.)
- [2] 北村智, 河井大介. 日記式調査法による情報行動の時間的連続性・断続性の検討—携帯インターネット利用とPCインターネット利用の比較分析—, 情報通信政策レビュー, 2014, Vol. 9, p. 114-143.
- [3] 橋内久美, 金丸隆太. 日記による「もつれた怒り」の低減, 茨城大学教育学部紀要, 教育科学, 2019, Vol. 68, p. 519-532.
- [4] 松田雛乃, 村井源. 日記文章の文体的特徴の抽出, 情報知識学会誌, 2020, Vol. 30, No. 2, p. 289-296.
- [5] 松田雛乃, 村井源. N-gramを用いた日記文章の文章構造の分析, 情報知識学会誌, 2021, Vol. 31, No. 2, p. 355-360.
- [6] 南野朋之, 鈴木泰裕, 藤木稔明, 奥村学. blogの自動収集と監視, 情報処理学会研究報告, 2004, Vol. 160, p. 129-136.
- [7] ムラウチドットコム. にほんブログ村, <https://blogmura.com/>, (参照 2021-8-22).