

## レビュー分析によるヒットするゲーム要素の抽出および評価手法の検討

メタデータ	言語: ja 出版者: Musashino University Smart Intelligence Center 公開日: 2024-03-25 キーワード (Ja): 自然言語処理, レビュー分析, スクレイピング, 生成AIの利活用 キーワード (En): 作成者: 白川, 桃子, 渡邊, 紀文, 宮田, 真宏 メールアドレス: 所属:
URL	<a href="https://mu.repo.nii.ac.jp/records/2000201">https://mu.repo.nii.ac.jp/records/2000201</a>

学生論文

## レビュー分析によるヒットする ゲーム要素の抽出および評価手法の検討

### Extraction of Key Elements for Hit Games by Review Data Analysis and Study of Evaluation Methods

白川桃子\* 渡邊紀文\*\* 宮田真宏\*\*

\*武蔵野大学 工学部数理工学科／副専攻 (AI 活用エキスパートコース)

\*\*武蔵野大学 教養教育リサーチセンター／MUSIC

#### 概要

現在、商品やサービスに対する評価にはレビューなどの自由記述が用いられる。近年の AI 技術の発達により自由記述の分析は自動化されているが、その手法はブラックボックスになっていた。この問題に対して本稿ではゲームソフトのレビューを対象に、レビュー内の単語毎に評価点をつけ、ヒットするゲームの特徴を抽出した。抽出したソフトの特徴を表す 6 つの単語を用い生成 AI によりゲームの説明文を生成し、既存ゲームの説明文と比較した。結果、ユーザーの属性ごとに説明文の解釈が異なること、および単語ごとに印象が異なるため、レビューの分析はユーザーの属性ごとに行うことが必要であるという結果を得た。

**キーワード：** 自然言語処理, レビュー分析, スクレイピング, 生成 AI の利活用

#### 1. はじめに

近年の科学技術の発展に伴い、従来データ化が困難であった情報もデータ化されアクセスできる時代となってきた。これらのデータは、人間が全体像を把握することが困難であるためビッグデータとも呼ばれている[1]。ビッグデータには売り上げ金額や販売数量などの定量化可能な数値情報だけでなく、自由記述や感情情報などのような定性的[2]な情報もある。現実世界において私たちが購買行動を起こす際に参考にする情報は、この定量的／定性的なデータの内、分析がしやすい定量的な情報だけでなく、定性的なデータにも多くの情報が含まれると言える。しかし、この定性的なデータを分析する為の絶対的な分析手法は現時点で試行錯誤段階である。一方で、半自動的に定性データを分析するツールとして UserLocal の AI テキストマイニング[3]や、DataRobot[4]などのツールもあるがそのデータの分析過程はブラックボックスであった。その仕組みを理解し使用するのには、有識者であればまだしも、初学者には困難であると言える。

本稿ではこの定性データ分析について、初学者が過程を容易に理解できる手法を提案し、その活用方法の例を示すことを目的とする。今回は、任天堂株式会社の販売している

Nintendo Switch[5]というゲーム機のソフトを対象とした。このゲーム機は国内外で2024年2月現在、計1億3,936万台以上[6]売り上げている。このNintendo Switchのソフトにおいても、販売数が伸びたものとそうで無いものは存在している。もしゲームの開発時に売れるゲームの傾向を知ることができれば、ソフトを開発する際の参考にできることが期待できる。そこで本稿ではソフトのヒット傾向は各ソフトのゲーム性が強く影響し、これを分析することでヒットするソフトに含まれる要素(単語)が明らかになると考えた。分析に使用するデータは、ユーザーの考えや感想を自由に記述でき、数値化したデータよりも具体的に表現されていると考えられるレビューデータを活用した。

さらに、抽出したヒットするソフトの単語を基にChatGPT3.5[7]を用いてゲームの説明文を生成し、既存のゲームの説明文と比較することで評価した。その際、ゲームへの関心が高いかどうかで分け、それぞれよりどのような評価が得られるかについても評価した。

## 2. データ収集, および前処理

### 2.1. 分析対象ソフトのリストの収集

全世界で人気となるソフトの特徴を抽出する為には国籍や年代を問わず、幅広いユーザーのレビューを収集することが必要であると考えた。そこで今回は、世界的に利用されているショッピングサイトの一つであるAmazon[8]を分析対象とし、Amazon内で販売されているソフトのレビューデータをスクレイピングツールの一つであるOctoparse[9]によりスクレイピングした(図1)。Octoparseにはセットアップが不要なプリセットテンプレートが複数用意されており、今回はその中から『[JP]キーワードスクレイピング\_Amazon』を使用し、まずは商品情報を収集した。このテンプレートでは、Keywordsに入力した単語を対象にAmazon内を検索し、検索結果の商品に関する情報を抽出する事ができる。今回は「switch, ソフト, 人気」の3つの単語を用いて商品情報を出力し、結果として75個の商品情報を収集した。



図1 Octoparseの実行画面の例

Figure 1 Example of Octoparse execution screen

## 2.2. ソフトのレビュー情報の収集、および前処理

ソフトの詳細なレビューの収集は、Octoparse のプリセットテンプレートより『[JP]レビュー詳細\_Amazon』を用いた。このテンプレートを利用するには Amazon 上の商品それぞれに割り当てられている 10 桁の番号を示す商品情報 (ASIN) [10]が必要となる。これは前節の結果に含まれるため、それを利用した。また、レビューは指定した期間に対して取得可能である。本稿では最新の流行を掴む事、および多くのソフトからレビューを収集することに重点を置き、2022 年 12 月 14 日～2023 年 6 月 14 日 (スクレイピング実行日から半年前まで) の期間を対象に計 13,324 件のレビューを取得した。

ここで取得されたレビューデータにはレビューのタイトルの有無や、役に立った人数など、複数の項目において欠損値が生じている場合がある。そこで本稿では収集したレビューデータ分析の前処理として、空欄のセルに対して該当する情報が文字列であれば「なし」、数値情報であれば「0」として補完した。

## 3. データ分析、および結果

### 3.1. 全レビューにおける頻出する名詞の抽出

それぞれのレビュー内に頻出する単語は、そのゲームを評価する重要な単語であると考えられる。そしてその中でも名詞は、ゲーム性を表現する際に重要な要素だろう。なお、名詞には「普通名詞、固有名詞、代名詞、数詞、形式名詞」の 5 種類があり、特に普通名詞は一般的なものの名前を示すという特性がある。これに基づき、全レビューの中から普通名詞を対象とした[11][12]。

分析は始めに、python で分析可能な形態素解析エンジンの一つである MeCab[13]を用いてレビューを分析した。ここでは、文章をあらかじめ設定した辞書情報に従い、形態素 (文章中の意味を持つ最小単位) に分解する。その後、各形態素に対して品詞や活用型、読みなどの情報を紐付けて出力する。なお、本稿ではレビュー内に含まれる多くの現代語に対応するためにシステム辞書を、mecab-ipadic-Neologd[14]に変更して利用した。これにより本稿では全レビューの中から名詞のみを抽出し、その上で各名詞の使用された回数のランキングを作成し、上位 100 語を重要名詞として以後の分析対象とした。

### 3.2. ネガポジ判定の実施

ユーザーがレビューを書き込む際にはユーザーの心的感情が含まれるため、レビュー内の単語に対しての評価は通常とは異なる独自の評価が必要となる可能性がある。本稿では、スクレイピングにより抽出されたレビュー中の 5 段階評価がこの単語の評価と関係すると考え、単語とレビュー評価の良し悪しを対応付けた。

本稿では、ポジティブな名詞は「名詞の登場回数が上位 100 位以内であり、かつ 5 段階評価中の評価 4 と 5 の合計割合が全体の 6 割以上の単語」とした (表 1)。表 1 にこの条件を満たした 26 個の単語のリストを示す。

自然言語処理では多くの場合、係り受けやその可視化[15]が行われる。この分析の目的の

表1 ポジティブな名詞一覧(全26単語)

Table 1 List of positive words (26 words in total)

ゲーム	購入	時間	気	自分	クリア
好き	子供	作品	私	今	そう
シリーズ	追加	レビュー	最高	版	世界
プレゼント	ソフト	満足	楽しみ	アクション	モード
今回	不満				

1つとして、頻出単語と共に使われる言葉を知ること、関連する重要語が抽出可能な点が挙げられる。そこで本稿では、抽出されたポジティブな26単語を含むレビューを対象に再度頻出単語を抽出した。結果、得られた単語の登場回数が1,000回以上であった単語を表2に示す。表2の結果は分析に用いたレビュー内の総単語数の内の約10.8%を占めるものであった。

表2の結果がゲームの構成要素に関連しそうな単語であると考えられるため、この中から「ストーリー、キャラ、時間、戦闘、クリア、子供」の6単語が人気の出るゲームに関連する単語であるとした。

### 3.3. レビュー中の頻出単語に基づいたソフトの内容の提案

ある要素を含むゲームを開発する際、その要素を持つゲームが本当に人気となるか事前に評価できれば、ゲーム開発における重要な観点となるだろう。そこで、前節までに得られた6つの単語を用いたゲームソフトの内容をOpenAI社の生成AIであるChatGPT3.5(2023-07-16に実行)に提案させた[7]。結果、「魔法の世界で盛り上がるストーリーに子供たち

表2 係り受け後の頻出単語とその出現回数の例  
(登場回数1000回以上制限付)

Table 2 Examples of frequently appearing words and their frequency of occurrence after engagement (with a limit of at least 1000 appearances)

単語	出現回数
ゲーム	3318
こと	2083
購入	2075
よう	2060
ストーリー	1830
方	1571
人	1420
これ	1261
キャラ	1209

単語	出現回数
時間	1195
気	1154
自分	1141
戦闘	1140
好き	1104
クリア	1104
それ	1086
的	1067
子供	1061

が夢中！個性豊かなキャラクターと一緒に冒険し、力を合わせて敵と戦いましょう。最後までクリアして、世界を救いましょう！」というゲーム内容が生成された（以下「提案ゲーム」）。

今回生成 AI により提案されたゲーム内容が人気ゲームの要素を含むならば、実際に販売されている既存の人気ソフトの評価と類似することが期待できる。そのため、2023 年上半期のダウンロードランキングが 1 位[16]であった「ゼルダの伝説 ティアーズ オブ ザ キングダム」[17]については、ChatGPT3.5 にゲームの概要を 50 文字程度で説明させた。結果、「プレイヤーは英雄となり王国を守り、謎解きと戦闘成長を遂げる。美しいグラフィックと感動的な冒険が待っている。」という説明文が生成された(以下「既存ゲーム」)。

#### 4. アンケートを用いた提案ゲームの評価

##### 4.1. アンケートによる回答者の属性と結果

ChatGPT3.5 により生成された 2 つのゲームの説明文が、ユーザーにどのように受け取られるかをアンケートにて評価した。本稿では、武蔵野大学工学部数理工学科や副専攻 AI 活用エキスパートコース 1 期生を中心とした計 160 人の結果を示す。本稿にてアンケートに回答していただいた方の性別毎の質問の回答を図 2 に示す。図 2 より回答者の約 77% が男性であった。

回答者のゲームへの関心度合いはゲームのプレイ頻度により異なり、これがゲームの評価にも影響すると考えられる為、アンケート結果よりゲームのプレイ頻度をグラフ化した(図 3)。図 3 より、ゲームのプレイ頻度が週に 3 日以内の人の割合は 53.1% であり、週に 4 日以上の人割合は 46.9% であった。さらに、24.4% の人がほとんどゲームを利用しない人であった。これを踏まえて本稿では、週に 4 日以上プレイするか否かで回答者の属性が 2 つに分けられると考え、プレイ日数が週 3 日以下の人を関心の低い人に、週 4 日以上の人を関心の高い人とした。

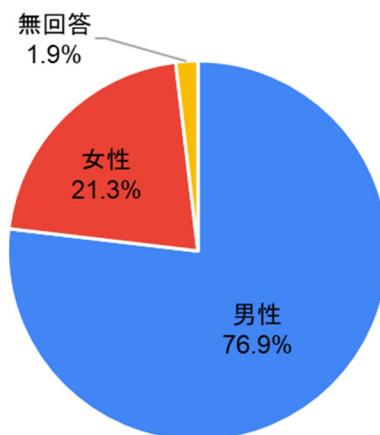


図 2 回答者の各性別の比率 (n=160)

Figure 2 Percentage of respondents by each gender (n=160)

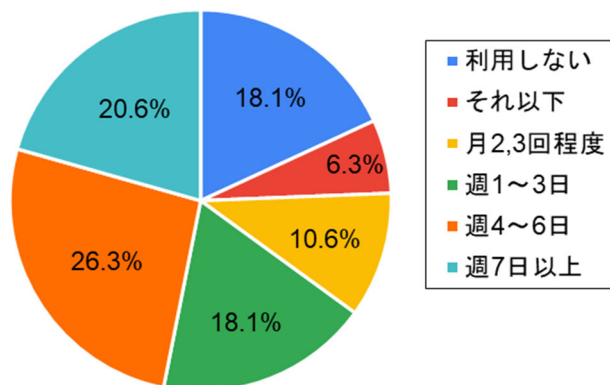


図3 回答者のゲーム機利用比率 (n=160)

Figure 3 Percentage of respondents using game consoles (n=160)

また、回答者の年代別にゲームの説明文の評価が異なる可能性がある。そのため、関心の低い人と関心の高い人毎にその属性に含まれる年代の人数をグラフ化した (図4)。図4より、それぞれの属性に関わらず、回答者の92%以上が20代以下であった。

それぞれの回答者は ChatGPT3.5 により生成された2つのゲームの説明文を見て、どの程度プレイしたいと感じたかを5段階 (プレイしたくない: 評価1~プレイしたい: 評価5) で評価した。図5に属性別の各ゲームの説明文の評価割合を示す。図5より、ゲームへの関心の低い人と高い人とでは、説明文を読んだ評価の割合が異なることが分かった。ゲームへの関心の低い人は、提案ゲームと既存ゲームの回答結果の割合の傾向が類似していた。さらに、低い評価 (評価1, 2) としている割合が44.7%程度であり、高い評価 (評価4, 5) としている割合が28.2%程度であったことから評価が低い結果が先行していた。一方で、ゲームへの関心の高い人の結果は、提案ゲームと既存ゲームとで回答割合の傾向が異なっ

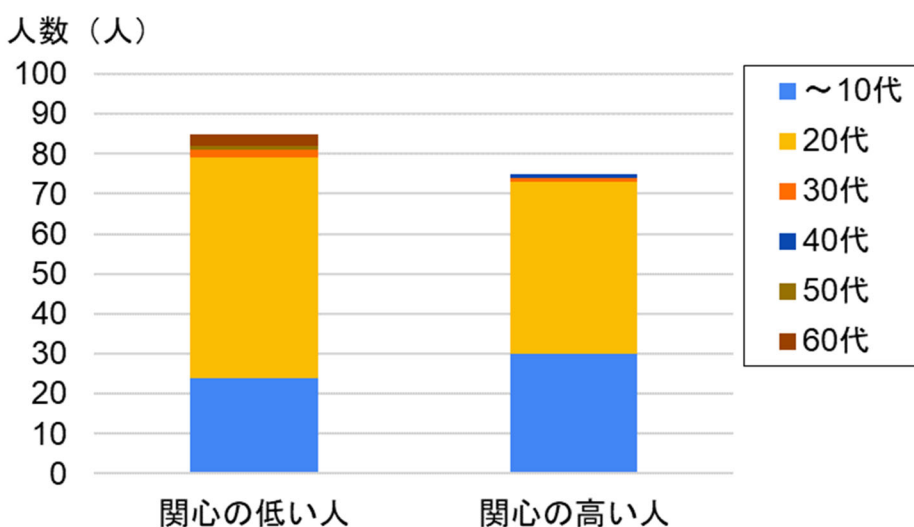


図4 ゲームの関心度別の回答者の年代の人数

Figure 4 Number of respondents of different ages by level of interest in the game

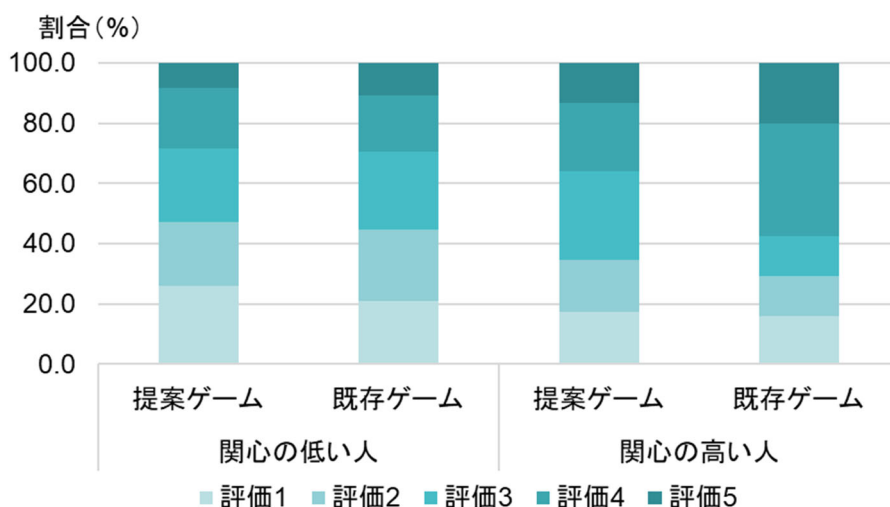


図5 ゲームの関心度別の  
各ゲームをプレイしたいと感じた割合

Figure 5 Percentage of respondents who felt willing to play each game  
by level of interest in the game

ていた。提案ゲームでは、評価1、2と回答した人の割合（34.7%）と評価3と回答していた人の割合（29.3%）と評価4、5と回答した人の割合（36.0%）が概ね類似する結果であった。これに対して既存ゲームでは、評価4、5と回答していた人の割合（57.3%）と評価1、2と回答していた人の割合（29.3%）が高い結果となった。

また回答者は各ゲームの評価の理由を自由回答にて回答した。その一部を以下に示す（表3）。表3より、提案ゲームか既存ゲームに関わらず、プレイしたい度が評価1～3では消極的な意見が、評価4、5では積極的な意見が多かった。評価1～3に着目すると、ゲームに対する興味が薄い、ゲームに興味はあるが何かしらのハードルを感じているといった、2種類の回答者が存在していた。評価4、5に着目すると、自由回答内で使用している言葉から、ゲームの特徴を掴み魅力を感じている人の意見が多くを占めている。また、提案ゲームの評価4、5には具体的なゲーム名も含まれていた。

#### 4.2. アンケート結果の考察

図5の結果より、ゲームに関心の低い人における評価1、2の割合が両ゲーム共に高かったことから、ゲームへの興味・関心の低さがゲームの説明文の評価にも影響していたと考えられる。また、提案ゲームと既存ゲームの評価傾向が同じであったことから、提案ゲームの説明文を生成するために抽出した6つの単語は、人気ゲームの要素を含んでいることを示唆する結果を得た。一方で、ゲームに関心の高い人における提案ゲームの評価が評価1、2と評価3および評価4、5の割合が概ね類似していたことから、好みにより判断が分かれる説明文であったと言える。これは今回ゲームの要素を取得する際にジャンル分けをせずに抽出したため、複数のゲームの要素を含む汎用的なゲームの説明であったことが原因として考えられる。これは、コメントに「没個性」というものがあつたことから判断できる。



表 3 提案及び既存のゲームに対する自由回答（抜粋）

Table 3 Free responses to proposed and existing games (excerpts)

ゲームの種類	プレイしたい度：評価 1～3	プレイしたい度：評価 4, 5
提案ゲーム	「子供たちが夢中」という謳い文句は自分には合っているとは思わない。	ゲームに非現実性を求めているので、魔法という点に強く惹かれたから。
	ゲームに興味が無いため。	ホグワーツレガシー[18]みたい。
	没個性	どんなストーリーなのかが気になってくるから。
既存ゲーム	あまりゲームに興味がない	グラフィック綺麗なの神
	もう少し若ければやってみたいのですが…	謎解きとアドベンチャーの組み合わせはやりこみがいがありそうで面白そうだから
	実際にゲームをプレイしてみないとプレイしたいかわからないから。	アクションゲームが好きだから。

この問題点を解決するためには、レビューに対して、ゲームのジャンルや流行っているものなどの情報を考慮した上で分析が必要であると考えられる。

また、表 3 のコメントにてプレイしたい度評価 1～3 の回答に「ゲームに興味が無いため。」「あまりゲームに興味がない」とあり、この人物のプレイ頻度が「利用しない」であった事から、ゲームへの関心の差が結果に影響したと考えられる。これは Amazon にレビューを書く人の多くはゲームが好きで書いていると考えられ、レビュー内特有の単語は普段ゲームをしない人にとってあまり聞きなれない単語であった可能性がある。

一方で、今回抽出した 6 つの単語を用いた説明文のコメントには「ホグワーツレガシーみたい。」とあるように実在するゲームと勘違いされる要素が含まれると解釈できる。これは、抽出した 6 つの単語をより我々が関連するゲームを考えるとファンタジー要素のあるゲーム（星のカービィ、キングダムハーツなど）が考えられるように、生成 AI においても、ファンタジー的な要素が適合すると判断されたことが原因として考えられる。

人がゲームの説明文を評価する際、単語自身が持つ印象が大きく影響すると思われる。具体的には、表 3 にて「ゲームに非現実性を求めているので、魔法という点に強く惹かれたから。」とコメントにあるように説明文内の単語がプレイ意欲を促進させるものがある一方で、「子供たちが夢中」という謳い文句は自分には合っているとは思わない。」というコメントにもあるように行動を抑制するものもあった。これらを踏まえると、対象となるユーザーの特性を考慮した上で使用する単語の抽出を行い、その上での評価が必要になるが、この点については今後の課題とする。

また本稿では、抽出した単語を用いて生成 AI により説明文を作成させ、アンケートにより評価したが、これは生成 AI を用いることで人の主観ではなく、様々な知識を踏まえたバランスの良い説明文が生成できると期待できるため、有用であると考えられる。

## 5. おわりに

本稿では従来ブラックボックスであった自由記述の分析方法を明確にすることを目的としてユーザーの自由記述であるレビューを分析した。ここではレビュー内で使われる単語とそのレビューの評価との関係を紐づけた独自の単語の評価を実施し、得られた 6 つの単語を用いて生成 AI によりゲームの説明文を生成した。その際、既存の人気ゲームについても同様に説明文を生成し比較対象とした。生成 AI を活用して説明文を作成することで、人の主観に寄らない説明文の生成をすることができたと考えられる。これらの説明文から得られるゲームの興味・関心をアンケートにより評価した。結果、ゲームの関心が低い人では、既存ゲームと提案ゲームに差はなかった。一方でゲームの関心が高い人では、提案ゲームの評価がいずれの条件においても類似する結果となった。これは、レビューを収集する際にジャンル等での制限を設けずにレビュー情報を収集したため、様々なジャンルを含むゲームの要素が抽出され、結果的に提案ゲームは汎用的なゲームの提案となったためであると考えられる。一方で、提案ゲームに関する自由回答にて実在する人気ゲームを想起した回答者も存在したため、ある一定レベルでの人気ゲームの要素は含まれていると言える。これはペルソナなどの手法を事前に活用しターゲットユーザーを明確にすることで、得られる結果も変わると考えられる。また、よりヒットするゲームを提案するためにはゲームのジャンルを絞ることや発売時期の新しいゲームに絞る、SNS などのデータを使用しリアルタイムな流行を把握するなど、いくつかの対応策が考えられるがこの点については今後の課題とする。

今後の展望は、本稿で示した手法と従来手法による結果との違いを評価することである。自然言語処理を用いた分析手法は多数存在するため、それらとの差分を示すことにより本稿で提案した手法の有用性をより示す可能性がある。

最後に、ゲームを構成するのはゲーム内容だけではなく、グラフィックや音楽、価格などの様々な要素がそのゲームをプレイしたいと思わせるか否かに関与してくる。これらの他の観点からもゲームソフトについて研究し、最適なゲームを提案し社会に還元したい。

## 謝辞

本稿の実現にあたり終始適切な助言を賜りました、武蔵野大学工学部数理工学科の佐々木多希子先生、加えてプログラム作成におきまして多大なるご協力いただきました平松勇紀さんに感謝の意を表します。

最後になりますが、本稿の評価のためにアンケートのご協力いただきました方々に、心よりお礼申し上げます。

## 参考文献

- [1] 総務省：“ビッグデータとは何か 第1部特集 ICTが導く震災復興・日本再生の道筋第1節 「スマート革命」-ICTのパラダイム転換-”，  
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc121410.html> (参照 2024-01-10)
- [2] UserLocal AI テキストマイニング：<https://textmining.userlocal.jp/> (参照 2024-03-09)
- [3] 株式会社エクス：“ビジネスに活用するデータ分析、定性データと定量データの違いとは？”，  
<https://digital-business.exa-corp.co.jp/blog/qualitative-and-quantitative.html> (参照 2024-01-10)
- [4] DataRobot：<https://www.datarobot.com/jp/> (参照 2024-02-14)
- [5] Nintendo Switch：<https://www.nintendo.co.jp/hardware/switch/index.html> (参照 2024-01-09)
- [6] 任天堂 2024年3月期 第3四半期決算短信〔日本基準〕(連結)：  
<https://www.nintendo.co.jp/ir/pdf/2024/240206.pdf> (参照 2024-02-09)
- [7] OpenAI ChatGPT：<https://chat.openai.com/> (参照 2022-12-29)
- [8] Amazon.com：[https://www.amazon.co.jp/ref=nav\\_logo](https://www.amazon.co.jp/ref=nav_logo) (参照 2024-02-09)
- [9] Octoparse スクレイピング|Web クローラー：<https://www.octoparse.jp/> (参照 2024-02-08)
- [10] アソシエイト・セントラル-アマゾン：  
<https://affiliate.amazon.co.jp/help/node/topic/G5KVDATAT5RKBBBG> (参照 2023-12-28)
- [11] 安武凌, 野中健一, 岩井将行(2020): “LexRank を用いた小説文章からの自動要約手法の検討”,  
 Web インテリジェンスとインタラクション研究会 予稿集 W12 研究会, pp.38-41
- [12] 山下皓太郎, 雲居玄道, 蓮本恭輔, 後藤正幸(2022): “BERT による特徴抽出を駆使した商品レビュー分析モデルに関する一考察”, 第 84 回全国大会講演論文集, Vol.1, pp.841-842
- [13] 京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所 共同ユニットプロジェクト GitHub: “MeCab: Yet Another Part-of-Speech and Morphological Analyzer”,  
<https://taku910.github.io/mecab/> (参照 2024-02-09)
- [14] 佐藤俊則 GitHub: “mecab-ipadic-neologd/README.ja.md at master”,  
<https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md> (参照 2024-02-09)
- [15] 株式会社リーディング・エッジ社：“自然言語処理入門 Vol.2 係り受け解析/構文解析”,  
<https://leadinge.co.jp/rd/2021/05/10/696/> (参照 2024-02-09)
- [16] Nintendo 2023 年上半期ダウンロードランキング：  
[https://store-jp.nintendo.com/download\\_ranking\\_2023.html](https://store-jp.nintendo.com/download_ranking_2023.html) (参照 2024-02-09)
- [17] ゼルダの伝説 ティアーズオブザキングダム|Nintendo Switch：  
<https://www.nintendo.co.jp/zelda/totk/index.html> (参照 2023-12-29)
- [18] ホグワーツ・レガシー-ホーム：  
<https://www.hogwartslegacy.com/ja-jp> (参照 2024-02-09)