

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

Supporting Exploratory Search in Comic Books Using Genre Information

山下 諒*¹
Ryo Yamashita

松下 光範*²
Mitsunori Matsushita

*¹関西大学大学院 総合情報学研究所
Graduate School of Informatics, Kansai University

*²関西大学 総合情報学部
Kansai University, Faculty of Informatics

Our research concerns the realization of an exploratory search system based on content information from comics. To realize such a system, it is necessary to obtain certain information, such as information related to the story and the characters. In previous research, we developed a system to extract information about comics using online reviews. In this paper, we propose a search system based on this extracted information. Further, in order to verify that this information is useful, we compare the proposed system to an existing comic search service. With the proposed system, more characteristics could be better understood using information obtained from reviews.

1. はじめに

日本における新刊コミックの年間出版数は年々増加しており、コミックを検索できるサービスも存在する。このようなサービスでは、書誌に関する情報 (e.g., コミックタイトル, 掲載誌) やコミックに付与された内容を表すタグ情報 (e.g., ラブストーリー, SF・ファンタジー) をクエリにして検索を行う。ユーザの情報要求が明確である場合は、これらの情報を用いて目的のコミックにアクセスできる。一方で、アクセスしたいコミックが曖昧なとき、例えば嗜好に合うコミックを探したいが自らの嗜好を明確に理解していない場合は、クエリを用いて検索しても満足できる情報にアクセスできるとは限らない。また、検索結果の各コミックに付与されているタグ情報を用いて繰り返し検索を行うことができるが、その場合単一のタグ情報のみを考慮した検索でしかなく、複数のタグ情報を加味した複雑な探索行為はできない。加えて、既存サービスで用いられているタグ情報の種類が少ないため、各コミックの内容における差別化が不十分であり、結果として膨大な数のコミックが提示される。このことが検索結果からコミックを選定するユーザにとって負担に成り得る。このような背景のもと、本研究では、自らの嗜好に合致したコミックへの円滑なアクセス支援を試みる。

タグ情報の種類が少ないことでユーザの要求に応じた検索支援ができていないという課題に対して、著者らはタグ情報に用いている情報を獲得するためにレビュー文に含まれる情報の抽出を試みた [9]。本稿では、ユーザの嗜好に合致したコミックへのアクセスを支援するための一検討としてレビュー文から抽出した情報を用いたコミック探索システムを提案する。提案システムでは、コミックタイトルをクエリとして、それと最も類似したコミックを判断しユーザに提示する。この時、各コミックにはレビュー文に含まれる各コミックの特徴を表す単語をタグ情報として付与している。さらに、タグ情報が複数共起しているコミックを類似コミックとして提示しており、複数タグを用いて類似したコミックにアクセスできる。また、レビュー文に含まれる情報がタグ情報として適しているのか検証するために、既存サービスで付与されているタグ情報を用いた場合と比較する。

2. 関連研究

2.1 コミックの情報アクセスに関する研究

コミックに関する情報は書誌に関する情報 (書誌情報) と内容に関する情報 (内容情報) に大別できる。コミックタイトルや掲載誌などの書誌情報は出版社や著者のホームページ、Wikipedia などから容易に獲得できる。一方で、コミックはテキストとイラストで構成されたマルチモーダルなコンテンツであるため、ストーリー情報や登場キャラクターの心情、人間関係などといった内容情報の抽出が困難である。

Morozumi らは、Wikipedia に記載されている各コミックの情報を抽出し、書誌情報や内容情報を FRBR (Functional Requirements for Bibliographic Records^{*1}) のモデルに基づいて整理している [5]。これにより、コミックに登場するキャラクター名や技名などの情報が獲得でき、そのコミックの特徴が判断できる。しかし、Wikipedia に記述されている情報の多くはコミック特有の固有名詞 (e.g., 登場キャラクター名, 主人公の通う学校名) であり、コミックを読む事で理解できる情報である。従って、未読のコミックの場合は、これらの情報を用いて内容情報を把握することは難しい。

岩間らは、Wikipedia に記載されている掲載誌名や作者名、テーマ (ジャンル) などのカテゴリに着目し、その情報を用いてコミックを探索的に検索できるシステムを構築している [13]。このシステムを用いることでカテゴリ単位でコミックを検索することは可能であるが、Wikipedia に記載されている各コミックのカテゴリの種類が少ないためコミック間の細かな類似性や相違性を判断することは難しい。これは、未読のコミックの中から嗜好に合致した作品を発見する際に用いるシステムとして適していないことを示している。

2.2 嗜好を考慮した情報アクセスに関する研究

本研究ではコミックを対象に、ユーザの嗜好に合致した作品への円滑な情報アクセス支援を試みているが、他のモダリティでも同様にユーザの嗜好を踏まえた情報アクセス支援に関する研究が精力的に進められている。

多田らは、各楽曲におけるアーティスト名やアルバム名の情報と楽曲内で用いられているメロディやハーモニー、リズムの

連絡先: 山下 諒, 関西大学大学院総合情報学研究所, 大阪府高槻市霊仙寺町 2-1-1, k809231@kansai-u.ac.jp

*1 <http://www.ifla.org/files/assets/cataloguing/frbr/frbr.pdf>
(3月18日存在確認)

時間的変化を表す情報の 2 種類の情報を楽曲特徴とし、それを用いて音楽を推薦するシステムを提案している [11]. このシステムでは、ランダムで再生される楽曲をスキップすることでその楽曲特徴に類似した楽曲は嗜好に合っていないと学習し、その楽曲特徴に類似していない楽曲を推薦する. この視聴行為とスキップ行為を繰り返すことでリスナーはその時の感性にあった楽曲を聴取することができる.

増田らは、書籍販売サイトに登録されている小説の中から嗜好に合致した作品へのアクセスを容易にするために、ユーザがどのようなクエリを用いているのか分析している [12]. その結果、“サスペンス”や“ファンタジー”などのジャンルを表す内容情報から“ほっこり”や“明るい”といったイメージなど、様々な表現をクエリとしていることを明らかにしている. これは、同じ娯楽メディアの 1 つであるコミックでも同様に様々な表現のクエリが用いられる可能性があり、本研究が目指すシステムでも多種多様なクエリの入力を可能にすることが望ましいと考えられる.

3. デザイン指針

本章では、本研究が目指すシステムを構築する際に、考慮しなければならない事柄について整理する.

3.1 クエリの入力に関して

システムを実現するためには、自身の要求をクエリとして入力する必要がある. しかし、人は必ずしも自らの趣味や嗜好を明確に認識しているとは限らず、外在化が難しい場合も存在する [10]. Taylor は、情報要求の明確さを「直感的な要求」「意識されている要求」「形式化された要求」「解決に至る要求」の 4 つに分類しており、上述してきた「曖昧な要求」ははじめの 2 つに該当する [7]. 曖昧な要求の場合は外在化が難しく、要求に沿った情報へのアクセスが困難であるため、クエリ生成の負担を軽減するようなデザインを検討する必要がある. 2.2 節でも述べたように、ユーザが自らの好みを入力する際には“サスペンス”や“SF”といった内容情報から“明るい”や“ほっこり”といったイメージまで、幅広いクエリが生成される可能性がある. 本稿で提案するシステムでは、入力クエリの負担の軽減に関する一検討として、入力クエリをユーザが好きな既読のコミックタイトルとする. ユーザが好きなコミックは嗜好に合致している作品であると判断でき、イメージなどの嗜好を直接入力するのに比べてクエリを生成しやすくと考えられる. 本稿で提案するシステムでは各コミックを特徴づける情報として著者らが分析してきたレビュー文に含まれる情報を用いる.

3.2 情報アクセスに関して

提案システムでは、ユーザの好きなコミックタイトルから嗜好の獲得を試みるが、必ずしもコミックに付与されている情報から正確なユーザの嗜好を抽出できるとは限らない. 例えば、検索時のユーザの心情 (e.g., 落ち込んでいる, テンションが高い) に応じて細かな嗜好が変化する可能性があり、それを読み取ることは容易ではない. ユーザの検索時の心情を考慮した情報アクセスを実現するために、提示する情報に対するユーザのフィードバックを用いる方法がある [6]. 繰り返し提示される情報の中からユーザに選定してもらうことで現時点のユーザの細かな嗜好を汲み取ることが可能になる. さらに、探索過程を通じてユーザは嗜好や要求を明らかにしていくこともできる. このような探索過程を繰り返しながら要求を明確にしていく情報探索モデルを、Exploratory Search と呼ぶ [8]. 情報探索の行為は Marchionini によって Lookup, Learn, Investigate の 3 種類に分けられており、3.1 節で述

べた「曖昧な要求」は Learn と Investigate に該当する [4]. Exploratory Search では、収束的検索 (Focused Searching) と発散的検索 (Exploratory Browsing) の 2 種類の検索パターンがあり、交互に繰り返しながら様々な情報に触れ、自らの要求を明らかにしていく. 発散的な情報検索を繰り返すことでより多くのコミックにアクセスできる機会が生まれるため、システムを用いることで新たなジャンルのコミックにもアクセスでき、ユーザにとって未知の嗜好に気づききっかけに成り得る.

3.3 情報提示に関して

ユーザが探索的に検索を行う際、多くのコミックを比較しながら嗜好に近いと思われる作品にアクセスしていくことで満足感を得る可能性はある. しかしながら、選択肢が多すぎると選択行為時に負担がかかってしまうため、提示する情報量を適切にしなければならない [3]. 加えて、コミック間の特徴を俯瞰できるように各々のコミックの特徴を提示する必要があるが、提示する情報が多すぎるとコミックを比較する際の負担に成り得るため、各コミックを特徴づける情報の提示量についても検討する必要がある.

コミックは、テキストとイラストで構成されたマルチモーダルコンテンツである. イラストの絵柄は作者に応じて大きく異なり、輪郭や線の太さ、目の描き方など様々な観点で絵柄の特徴が見られる. コミックを選定する際、コミックの内容だけでなくこの絵柄の好みも嗜好の合致の判断に影響すると考えられる. 提案システムでは、ユーザが嗜好に合致するコミックを探す際にイラスト情報も判断の基準にできるようにコミックの絵柄に関する情報も提示することが望ましいと考える.

4. 実装

4.1 提案システムに用いる情報

本稿で提案するシステムの実装に必要な情報を得るためにコミック専用のレビューサイトである漫画レビュー.com^{*2}を用いた. また、比較を行うための情報源として電子書籍販売サイトであるコミックシーモア^{*3}を対象とした. 今回扱う情報は、漫画レビュー.com とコミックシーモア共に登録されているコミックかつ漫画レビュー.com 内で 20 以上のレビューが付与されている人気コミック上位 100 作品とした. 漫画レビュー.com から抽出した情報は、テキスト形式であり、言語処理技術を用いて各レビュー文の特徴を抽出する必要がある. 本稿では、各コミックの特徴を獲得するために 100 作品のレビュー文に対して TF-IDF 法を用いた. TF-IDF 法とは、文書に含まれる単語の相対的な重要性を表す指標であり、単語の出現頻度を表す TF (Term Frequency) 値と文書頻度を表す DF (Document Frequency) 値の逆数の積で表される. TF-IDF 法を用いることで、各々のレビュー文に記載されている特徴的な単語を確認することができる. 今回は、算出結果のうち TF-IDF 値の上位 50 語 (e.g., “寄生”, “パラサイト”, “グロテスク”) を各コミックを特徴づける情報とした. 一方で、コミックシーモアから抽出した情報はサイト内で各コミックに付与されているジャンル情報 (e.g., “SF・ファンタジー”, “青年マンガ”, “ホラー”) とした. コミックシーモアに登録されているタグ情報が名詞情報に限られているため TF-IDF 法を用いた特徴量算出も名詞を対象とした. なお、漫画レビュー.com に投稿されているレビューの数は各コミック異なっており特徴量の算出時に影響を及ぼす可能性がある. そこで、本稿では、各コミックの TF 値を各コミックのレビュー数で除算し正規化した [9].

*2 <http://www.manngareview.com/> (3月18日存在確認)

*3 <http://www.cmoa.jp/> (3月18日存在確認)

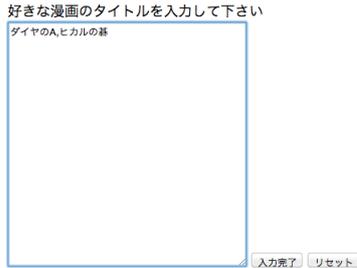


図 1: 提案システムの初期画面 (入力画面)

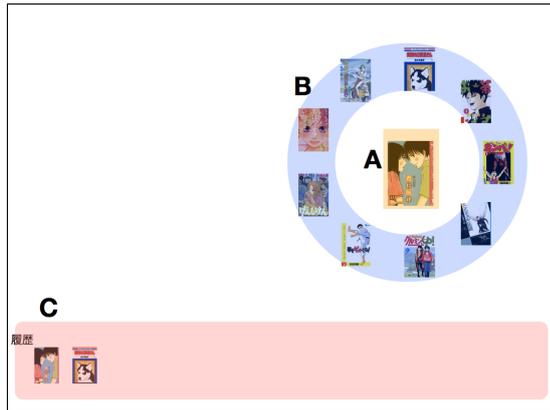


図 2: コミック探索画面

4.2 提案システムの全体像

本稿で提案するシステムを図 1, 図 2 に示す。システムのインタフェース部分は HTML と JavaScript で作成しており、ジャンル情報の集約と類似度の計算には Python を用いた。なお、今回の類似度の計算にはコサイン類似度を用いている。システムを利用するユーザは、図 1 のテキストエリア部分に好きなコミックタイトルを入力し、“入力完了”のボタンをクリックする。その後、図 2 の画面に遷移し、図 2-A に入力されたコミックと最も類似度が高いコミックが提示される。各コミックの表紙情報を用いているため嗜好の合致の判断の際、絵柄も考慮して行うことが可能である。図 2-A のコミックの周囲に展開している図 2-B には図 2-A のコミックと類似したコミックが提示されており、類似コミックを選択することでそのコミックが中央に遷移し、またその周囲には選択したコミックの類似コミックが提示される。この類似判定には、各コミックにタグ付けしたジャンル情報の共起数を用いており、提案システムでは 2 つ以上のジャンル情報が共起したコミックを類似コミックとしている。この類似コミックが膨大に提示された場合、3.3 節で述べたようにユーザが類似コミックの中から選択する際に負担がかかる。今回は、類似コミックが 10 作品を超える場合はその中からランダムに 10 作品を選択して提示した。各コミックの表紙をマウスオーバーすることでコミックにタグ付けされているジャンル情報を閲覧することができる。また、中央に表示されているコミックは選択することによって図 2-C の部分に蓄積されるため、最終的にユーザがコミックを選定する際の参考に成り得る。

5. 議論

5.1 本稿での到達点

提案システムでは各コミックにタグ付けする情報として、Web 上のレビュー文に含まれる情報を用いた。図 3 にその情報の一例を示す。また、図 4 にコミックシーモアで各コミックにタグ付けされている情報を用いた例を示す。これらの結果を比較するとレビュー文から抽出した各コミックの特徴の方が既存サービスでタグ付けされている情報よりも詳細であることがわかる。従って、コミックの特徴をより把握することができるため、よりコミックの内容を直感的に判断することが可能になり、嗜好の合致の判断が容易になったと思われる。

図 4 で、既存サービスのタグ情報を用いた探索過程を表しているが、この図で示されている類似コミックはランダムで選別されたものであり、実際にはこの作品数を遙かに凌ぐ。これは、既存サービスで付与されているタグ情報では、各コミックを十分に差別化できていないことが分かる。一方、提案システムでは対象とした 100 作品のうちスポーツ関連のコミックが多く、一部で 10 作品を超える作品が類似コミックであると判断される場面も見受けられたが、概ね各コミックの類似コミックが数作品であったことから、既存サービスよりも各コミックを差別化できていると言える。

5.2 提案システムにおける課題

今回のレビュー文からの情報抽出には、TF-IDF 法を用いた。また、特徴量算出を行うために MeCab^{*4} を用いて形態素解析を行った。形態素解析を行うことで名詞情報の判別が自動で行えるが、2.1 節でも述べた未読のコミックにおいて内容を判断することが難しい登場キャラクター名やその単語単体でコミックの内容の判断が困難な情報が混在する。登場キャラクター名に関しては、Wikipedia に記載されている各コミックの“登場人物”の項目に含まれる情報から抽出し、ストップワードとして設定することで解決できると考えられる。一方で、内容の判断が困難な単語 (e.g., 書誌に関する情報、形態素解析の誤分割) は、今回の手法で単語の意味まで考慮した分析を行うことが難しいため、それらの単語を判別して除外することは容易ではない。著者らによる先行研究でレビュー文に含まれるテキストの意味を考慮した分析を行うために TF-IDF 法に加えて、Latent Dirichlet Allocation (LDA) を用いたトピック分析を行った [2, 9]。LDA を用いることで感想などのレビュー文中に含まれる単語の意味にまで考慮したトピック推定を行うことができるため、よりレビュー者の意見を反映したコミックの内容情報が獲得できると期待していた。しかしながら、LDA は教師なし学習であり、各レビュー文中に含まれる曖昧なニュアンスを考慮したトピック推定ができず、各トピックの特徴を判断することができなかった。この結果を踏まえて、LDA の発展的技術の 1 つである Hierarchical Latent Dirichlet Allocation (hLDA) を用いた分析を検討している [1]。hLDA はトピック間の関係性を階層的に推定して構築することができるため、階層関係を用いてトピックの関係性やコミック間の関係性の俯瞰が可能になる。レビュー者がレビューを記述する際に暗黙的に省略している情報は存在すると考える。例えば、サッカーがテーマのコミックに対してレビュー者はサッカーという単語は多用せずに、登場人物に関する感想や好みのシーンを記述すると思われる。hLDA を用いることによってレビュー者が発信する情報がどの階層の話であるのか推定することができ、LDA の推定結果に比べて良い結果が得られると期待できる。

*4 <http://www.mecab.sourceforge.net> (3 月 18 日存在確認)



図 3: レビュー文から抽出した情報を用いたタグ情報 (拡大図)



図 4: 既存サービスの情報を用いたタグ情報 (拡大図)

5.3 今後の展望

今後の展望として、まずシステムで探索可能なコミックの数を増やす必要があると考える。今回は試作として 100 作品を対象に探索可能なシステムを構築したが、探索過程を繰り返しながらユーザの要求を明らかにするためには、より多くのコミックの情報に触れる方が望ましい。今後は、対象作品を増やして実際にユーザ観察を行い、コミック探索に適したインタフェースへの改良を行っていく。また、履歴部分に蓄積されるコミックはユーザが注目して選択したコミックであるため、それらの情報はユーザの嗜好に近い可能性が高い。今後は、履歴部分に蓄積された情報を用いて図 2-B で提示される類似コミックが変化するように各ユーザに適した情報を提示する機能の構築も行っていく。

6. おわりに

本稿では、コミックの内容情報を考慮した探索的検索システムの実現に向けた一検討として、Web 上のレビュー文に含まれる情報を用いた探索システムを提案した。各コミックの特徴量を算出するために、TF-IDF 法を用いて、レビュー文から特徴語の抽出を行い、TF-IDF 値上位 50 語を今回対象とした 100 作品にタグ付けた。また、レビュー文から抽出した情報がコミックの探索過程において適した情報であるのかを検討するために既存のコミック検索サービスで用いられているタグ情報を用いた場合と比較した。その結果、レビュー文から抽出した情報の方がそのコミックの特徴をより詳細に表していることがわかった。一方で、今回用いた手法ではレビュー文中のテキストの意味を考慮した分析は行っていない。今後は階層的な

トピックの分析が可能な hLDA を用いて、階層モデルに基づいてテキストの意味を考慮した分析を行っていきたい。

参考文献

- [1] David. M. Blei, Thomas. L. Griffiths, and Michael. I. Jordan. The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies. *Journal of the ACM*, Vol. 57(2), No. 7, 2010.
- [2] David. M. Blei, Andrew. Y. Ng, and Michael. I. Jordan. Latent Dirichlet Allocation. In *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.
- [3] Sheena S. Iyenger and Mark R. Lepper. When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology*, Vol. 79, No. 6, pp. 995–1996, 2000.
- [4] Gary Marchionini. Exploratory Search: From Finding to Understanding. In *Communication of the ACM*, Vol. 49, No. 4, pp. 41–46, 2006.
- [5] Ayako Morozumi, Satomi Nomura, Mitsuharu Nagamori, and Shigeo Sugimoto. Metadata Framework for Manga: A Multi-paradigm Metadata Description Framework for Digital Comics. In *Proceedings of International Conference on Dublin Core and Metadata Application*, pp. 51–60, 2009.
- [6] J. J. Rocchio. Relevance feedback in information retrieval. *The Smart Retrieval System: Experiments in Automatic Document Processing*, pp. 313–323, Prentice-Hall Inc., 1971.
- [7] Robert S. Taylor. Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries*, Vol. 29, No. 3, pp. 178–194, 1968.
- [8] Ryen W. White. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan and Claypool Publishers, 2009.
- [9] Ryo Yamashita and Mitsunori Matsushita. Content discrimination of comics based on users' reviews. In *The Third Asian Conference of Information Systems*, pp. 79–85, 2014.
- [10] 土方嘉徳. 推薦システムにおけるインタラクション研究へのいざない. *ヒューマンインターフェース学会誌*, Vol. 15, No. 2, pp. 131–134, 2013.
- [11] 多田圭吾, 山西良典, 加藤昇平. ユーザ感性へのインタラクティブ適応に基づく楽曲推薦システム. *情報科学技術フォーラム講演論文集*, Vol. 11, No.2, pp. 23–29, 2012.
- [12] 増田純太, 杉本徹. 小説推薦システムの構築に向けた検索表現と書評の分析. *情報科学技術フォーラム講演論文集*, Vol. 11, No.2, pp. 189–190, 2012.
- [13] 岩間勇介, 三原鉄也, 永森光晴, 杉本重雄. オントロジーと LOD に基づくマンガ配列の可視化による探索支援システム. *HCG シンポジウム 2014 論文集*, pp. 357–361, 2014.