

ワードクラウドを用いた選書支援手法とその実装

黒柳 竣¹ 後藤 大輝¹ 濱川 礼¹

概要: 本論文では、評価の高いレビューとレビュー内にある引用文章を動画形式のワードクラウドで表示することで選書支援を行うことを目的とした手法とその実装について述べる。近年、新型コロナウイルス感染拡大の影響もあり、屋外活動の機会は減少し、読書の機会が増加した。それに伴い、書籍を選ぶ機会も増加したが、タイトル、著者名、帯や表紙の情報から好みの一冊を見つけることは困難である。そこで、有益な情報となるのが、読者の意見をまとめたレビューサイトであるが、中には的外れな意見もあるため、必ずしもこれから書籍を読む上で参考になる情報ばかりとは言えない。そこで、本研究では、レビューの中に含まれる書籍の引用文と評価の高いレビューをその文章の頻出度合いやレビューの評価を指標に文章の大きさを変化させ、動画形式のワードクラウドで表示することで、レビューを読む手間を減らし、さらに書籍内の文章という書籍を選ぶ上での新しい評価基準を提示することで選書支援を行う。

A word cloud-based book selection support method and its implementation

Abstract: This paper describes a method and its implementation that aims to support book selection by displaying a word cloud of highly rated reviews and citations in the reviews in a video format. In recent years, opportunities for outdoor activities have decreased and opportunities for reading have increased, partly due to the spread of the novel coronavirus. This has increased the number of opportunities to select books, but it is difficult to find a book of one's choice based on the title, author's name, obi, or cover. However, some of the opinions may be off the mark, so the information may not always be helpful in reading a book. Therefore, in this study, we propose a new evaluation criterion for selecting a book by displaying the book quotations and highly rated reviews included in the reviews in a word cloud in animation format, changing the size of the sentences based on the frequency of occurrence of the sentences and the rating of the reviews, thereby reducing the time required to read the reviews. This system also supports book selection by presenting a new evaluation criterion for selecting books, namely, the text within the book.

1. はじめに

近年、新型コロナウイルスの感染拡大の影響を受け、屋外活動の機会は減少し、読書などの屋内で活動する機会が増加した。また、出版科学研究所によると、2021年の書籍の売上は前年比3.6%増の1兆6742億円となっており、3年連続で増加している(図1)[1]。このことから新型コロナウイルスが感染拡大する以前より読書の機会が増えたことがわかる。読書をする機会が増えたことで、書籍を選ぶ機会も必然的に増えた。しかし、現在、実際に本屋に行って好みの1冊を選ぶ際に与えられている情報は、多くの場合、タイトル、著者名、帯や表紙の情報のみであり、これらの情報から書籍の内容をくみ取るのは非情に難しい。ま

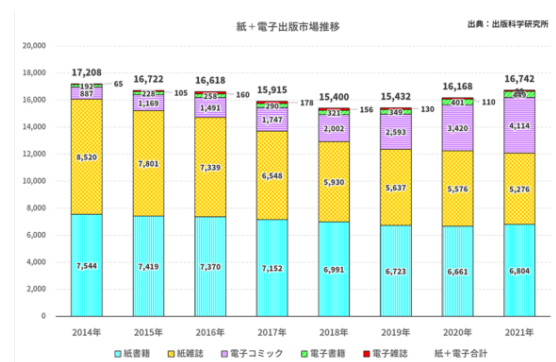


図1 紙+電子出版市場の推移

た、アニメや映画などと比べて、書籍は、あまり大々的にCMなどで宣伝が行われないため、なおさら読者が書籍の内容を掴むのは難しい。そこで、有益な情報となりうるの

¹ 中京大学工学部情報工学科

は、読者メーター [2]、ブックログ [3]、読書ログ [4] などの読者の意見をまとめたレビューサイトである。これらのサイトは現在、多数存在しており、多くの読者が自分の読んだ書籍について感想を書き込んでいる。基本的に、読者のレビューは、作品の印象を把握するためには、最適な情報だといえるが、レビューの中には、的外れなことを言うような意見や「もっとこうしたほうがいい」といったような作者に対しての意見なども混ざっており、必ずしもこれから読む上で参考になる情報ばかりとは言えない。また、稀にネタバレ要素を含むレビューが存在するため、ネタバレを隠す機能などで対処をしているレビューサイトもあるが、ネタバレかどうかを伝えるのは、ユーザに任されてしまっているため、自分のレビューがネタバレであることを伝えずにレビューを書き込むユーザも少なくない。また、レビューは、書籍 1 冊ごとに数多く書かれており、それらの文章を毎回読んで書籍を選ぶのは大変である。そこで、本稿では、多くの読者の共感を得たと考えられる評価の高いレビューとレビュー内にある作者特有の文体を表した引用文章をワードクラウドを用いて図 2 のように視覚的にわかりやすく表示し、より既存のレビューを読みやすくすることで書籍の選定支援を行う。ネタバレを含む文章は、見ないようにする人が多いため、評価が上がりにくい。そのため、評価の高いレビューは、ネタバレを含みにくいと考える。引用文も同様にわざわざネタバレを含む文章を引用する人は少ないため、ネタバレを含みにくいと考える。そのため、評価の高いレビューや作者の引用文を提示することで、意図せずネタバレを含むレビューを見てしまうという問題も解消できると考える。

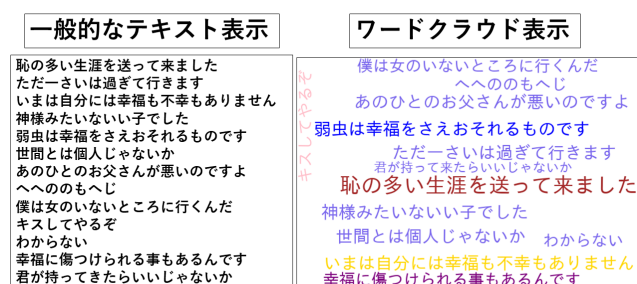


図 2 一般的なテキスト表示とワードクラウド表示の比較

2. 関連研究

本章では、ワードクラウドに関する研究及び選書支援に関する研究について述べる。

2.1 選書支援に関する研究

藤田らは文章をネガポジ判定し、その結果をもとにした

文章の感情の起伏を折れ線グラフを表示することで、小説の印象の推移を可視化する研究を行った [5]。また、村井らは小説の本文とレビューを用いて、物語の進行に合わせて、読者の興味を惹いた単語を可視化することで、立ち読みを支援する研究を行った [6]。これらの研究は、文章の流れや読者の興味を惹いた単語をユーザに提示することで、選書支援を行った。しかし、これらの研究では、文章の詳細な情報や文章のスタイルがわからず、書籍を選ぶ上での情報が不足していると考える。そのため、本稿では、評価の高いレビューや作者の文体がわかる引用文をユーザに提示する。

2.2 ワードクラウドに関する研究

評価の高いレビューや引用文をユーザに提示する際はワードクラウドを用いて表示を行う。ワードクラウドに関する研究は、多くなされており、毛利らはより短時間で商品の意見を把握するため、商品レビューで書かれているポジティブな意見とネガティブな意見を極性辞書によって判別し、それらの意見を表示する際にワードクラウドを使用した [7]。また、上野らは授業中の掲示板書き込み内容を可視化する際にワードクラウドを用いた [8]。これらの研究で、情報理解の支援としてのワードクラウドの有用性が示されている。そのため、本稿でも、同様にワードクラウドを用いる。しかしながら、従来のワードクラウドは、品詞ごとで文章が区切られて表示されるため「良い」、「悪い」、など単純な情報を理解するには最適であるが、「どのようにいい」、「どのように悪い」といった詳細な情報を理解するには、あまり適切ではない。そのため、本稿では、文章を品詞で区切らずに、文章のまま表示を行う。文章のまま表示する際に、情報量と見栄えが天秤にかけられてしまうという問題があるが、ワードクラウドを画像形式でなく、動画形式で表示することで、その問題に対して改善を行う。

3. 提案システム

提案システムは Python を用いて開発を行った。本システムは書籍の検索・検索結果表示、レビューのスクレイピング、文章の判別、ワードクラウド作成、ワードクラウド表示の 5 つの工程によって構成されている (図 3)。以下でそれぞれの工程について詳しく述べる。

3.1 書籍の検索・検索結果表示

ユーザは、タイトルや著者名、出版社名、ISBN コードを入力し、検索を行う (図 4)。検索結果は図 5 のように表示される。

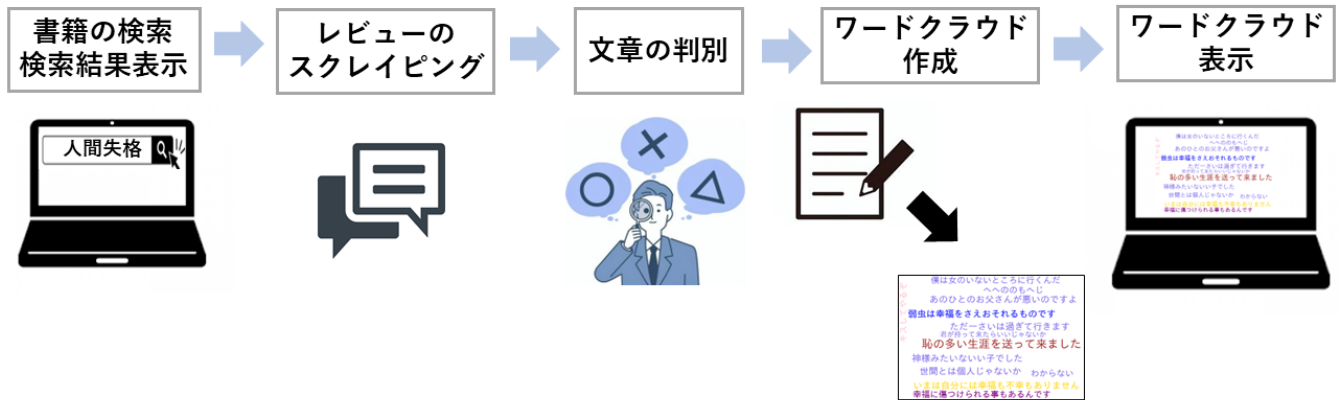


図 3 システムの流れ



図 4 検索画面



図 5 検索結果表示画面

3.2 レビューのスクレイピング

上記の図5の本の詳細ボタンを押すことで、ワードクラウドの作成を行う。ブックウォーカーが運営する日本最大級の読書コミュニティサイト「読書メーター」(図6)のレビューをスクレイピングし、レビューの抽出を行う。レビューとともにそのレビューに対する評価数も一緒に抽出

を行う。

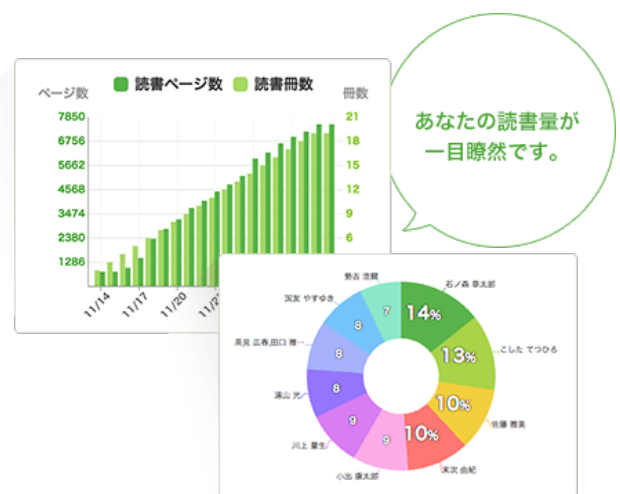


図 6 読書メーター

3.3 文章の判別

抽出されたレビューの中から、書籍の引用文と思われる文章とそうでない文章との判別を行う。書籍内の文章と思われる文章であるかどうかの判断基準は以下の2つとする。

- (1) 「」や『』で囲まれている
- (2) 名詞ではない

この判断基準に沿った文章を書籍の引用文と仮定し、判断基準に沿わない文章を書籍の感想を述べている文章とす

る。また、この際に、書籍の引用文の評価数は、書籍のレビュー中の引用文の頻度とする。

3.4 ワードクラウド作成

書籍の引用文と書籍の感想を述べている文章の判別を行った後、それぞれの文章を入力として、書籍の引用文を表示するワードクラウドと書籍の感想文を表示するワードクラウドの2つのワードクラウドの作成を行う。ワードクラウドを作成するうえでの主な設定としては、以下に記述する。

- 表示する文章
評価数が高い順に上から14つ選出したものとする。
- 各文章の大きさ
レビューの評価数の高さに応じて、文章の大きさを变化させる。
- 各文章の長さ
文章があまりに長い場合、画像内に文章を収めるために、文章のサイズが小さくなってしまったため、文章の文字数の上限は25に制限した。
- 各文章の色
各文章に対して、PythonのライブラリであるML-Ask[9]を用いて感情分析を行い、その結果をもとにブルチックの感情の輪[10]を参考にして図7の配色で色を变化させた。

嫌	■
喜	■
怖	■
安	■
好	■
哀	■
怒	■
驚	■
昂	■
無	■

図7 各感情の配色

また、本システムでは、複数のワードクラウドの画像をパ

ラパラ漫画のようにすることで、動画形式にして表示を行う。また、画像を变化させる間隔は1.5秒に設定しており、1.5秒ごとに文章が追加されていくように表示を行った。付録でワードクラウドの構成画像の例を記載した。

3.5 ワードクラウド表示

本システムでは、書籍の基本的な情報(タイトル、著者名、ISBNコード)などとともにワードクラウドを表示させる。また、2つのワードクラウドは、楽天でのレビューの下にある変更ボタンを押すことで、下記の図8と図9のように变化させることが可能となっている。

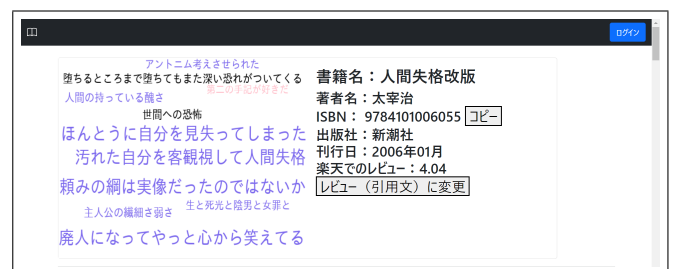


図8 感想文を入力としたワードクラウド表示画面

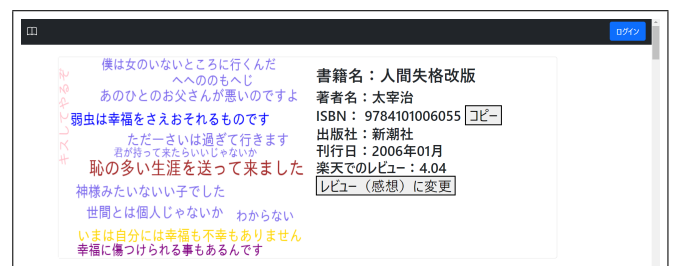


図9 引用文を入力としたワードクラウド表示画面

下記の図10のようにワードクラウドの下部分にワードクラウドの入力元となった文章も表示することで、ワードクラウド内で表示される文章に限度があるという問題への対処を行っている。また、書籍の中には、レビューが少ない書籍もあるため、ワードクラウドで表示される文章の質が悪かったり、まず、表示される文章が全くないという問題が生じるため、文章の投稿、評価という機能の実装も行った。各文章の横にある評価ボタンが押されることで、文章の評価数が増える仕組みとなっており、それに伴って質の悪い文章が排除され、質の良い文章がワードクラウド内に表示されることが期待される。また、文章の投稿が行われることで、入力元のレビューの少なさを補うことができると考える。特に、書籍の引用文は、あまり量が多くないため、文章の投稿機能でこの問題が解決することが期待される。

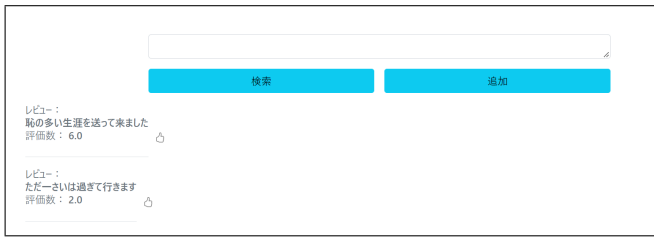


図 10 入力元文章表示画面

4. 実験

本章では、本システムで作成されたワードクラウドの有用性を明らかにするために行った被験者実験について述べる。

4.1 ワードクラウドの有用性評価

現状、本システムは、ローカルネットワークでしか、使用することができないため、多くのユーザに使っていただくことで機能することが想定される文章の投稿、評価の機能が全く意味をなさない。そのため、書籍によっては、抽出した引用文のなかに書籍内の文章ではないものが生じてしまうことがある。そこで、本実験では、多くのユーザに文章の投稿、評価を行ってもらい、書籍内の文章のみがワードクラウド上に表示されるという状況を想定して行った。具体的には、文章が一般に公開されている青空文庫内の作品を対象とすることで、引用文の中から、書籍内の文章ではないものを削除し、引用文のワードクラウドの作成を行った。

4.2 実験方法

大学生 10 人に対して、ワードクラウドの有用性を明らかにするための実験を行い、その後アンケート調査を行った。4 個の 5 段階評価設問、自由記述式設問を用意した。対象とした作品は、青空文庫の 10 作品である (表 1)。

表 1 対象の作品

タイトル	作者
吾輩は猫である こころ 三四郎 夢十夜	夏目漱石
斜陽 人間失格 走れメロス	太宰治
墮落論	坂口安吾
学問のすすめ	福沢諭吉
銀河鉄道の夜	宮沢賢治

以下に、評価を行ってもらったワードクラウドの一部を示す (図 11, 図 12, 図 13, 図 14)。

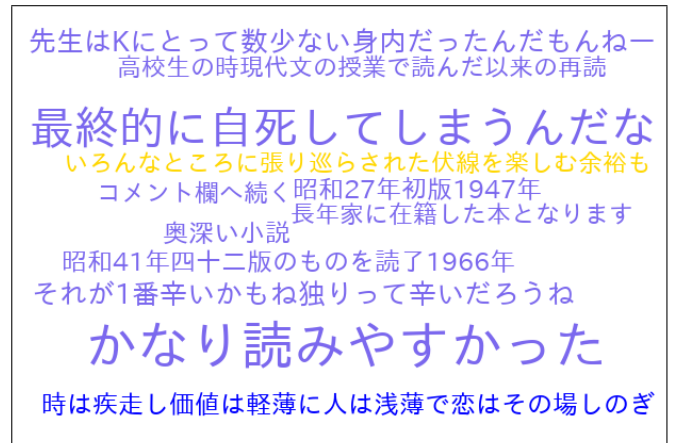


図 11 感想文を入力としたワードクラウド (こころ)

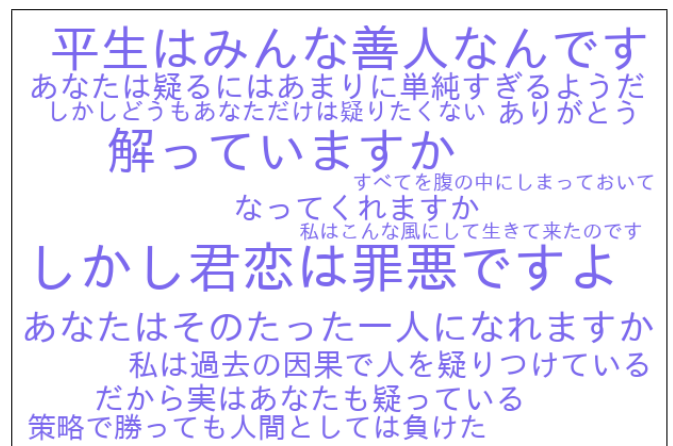


図 12 引用文を入力としたワードクラウド (こころ)

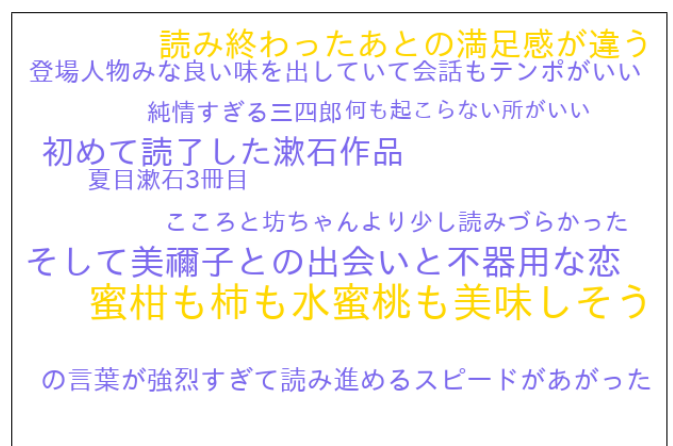


図 13 感想文を入力としたワードクラウド (三四郎)

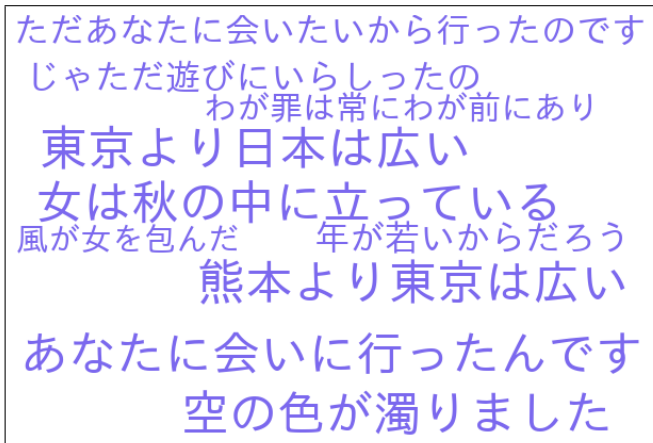


図 14 引用文を入力としたワードクラウド (三四郎)

4.3 結果と考察

アンケートの設問とその結果を表 2, 3 に示す. 設問に対して 5 段階評価を用意し, 1 が「いけない」～5 が「いえる」として回答してもらった. また上記の設問ごとに任意でコメントを自由記述で書いてもらった. なお, 平均値は小数第二位で四捨五入を行っている.

4.3.1 読んだことがある作品に対するの評価

10 個の作品の中から読んだことがある作品がある被験者に対して, 行ったアンケートについて以下に示す.

- 作品を的確に表したレビュー (感想) が表示されているといえるか

上記の質問に対して, 結果の平均が 4.0 点と高い数値を示した. 意見としては, 「同感できるレビューが見受けられたため」, 「身に覚えのある単語がおおい」などが挙げられた. しかし, 「一部のレビューは作品の内容とは関係ないものもあったから」, 「内容とは関係ない部分の感想が抜かれることがある」などの否定的な意見も挙げられた.

- 作品を的確に表したレビュー (作中の文章) が表示されているといえるか

上記の質問に対して, 結果の平均が 4.4 点と高い数値を示した. 意見としては, 「覚えのあるフレーズが沢山あったため」, 「有名な台詞が見られたから」などの評価の高い意見が多く挙げられたため, 引用文を入力としたワードクラウドの質は概ね高かったことがわかる.

4.3.2 読んだことがない作品に対するの評価

10 個の作品の中から読んだことがない作品がある被験者に対して, 行ったアンケートについて以下に示す.

- レビュー (感想) のワードクラウドは本を選ぶ上で有効

だといえるか

上記の質問に対して, 結果の平均が 4.3 点と高い数値を示した. 意見としては, 「色々な意見を見て作品に対して興味がわく」, 「人間ばい感想があれば, 参考になると思った」などが挙げられた. しかし, 「個人的に少し情報量として足りないと感じた」などの否定的な意見も挙げられた. 本実験では, 本システムで作成されたワードクラウドのみを評価してもらったが, 実際にユーザに使ってもらう際は, ワードクラウドとともに入力元のレビューも表示を行うため, この問題は解消できると考える.

- レビュー (作中の文章) のワードクラウドは本を選ぶ上で有効だといえるか

上記の質問に対して, 結果の平均が 3.6 点とレビュー (感想) のワードクラウドに関する質問と比べると少し低い数値となった. 意見としては, 「気になるワードがいくつかあって惹かれたから」などの評価の高い意見もあったが, 「感想の方と比較すると弱い感じがした」, 「あらずじくらは何となく推測できるが, どういったジャンルなのかがわかりにくかった」といったレビュー (感想) のワードクラウドと比べるとあまり有効ではないという意見が挙げられた.

4.3.3 全体を通しての評価

アンケートの最後に, 本システムで作成されたワードクラウドを見るのと, レビューサイトのレビューを見るのでは, どちらが本を選ぶのに有益であるかという質問を行った. 結果を以下に示す (図 15).

普通のレビューを見るのと, 生成されたワードクラウドを見るのを比べた場合, どちらが本を選ぶのに有益であるか?

10 件の回答

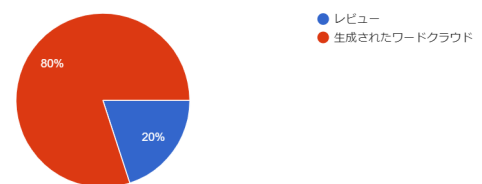


図 15 全体を通しての評価結果

多くの人がレビューサイトのレビューを見るよりも本システムで作成されたワードクラウドを見るほうが, 本を選ぶ上で有益であると答えており, このことから, このシステムの有用性がわかる. また, 全体的な改善点として挙げられた意見としては, 「1つ1つの文章の出るタイミングが早い」, 「文章の出し方をもっと工夫できたらなお良い」等の意見が挙げられた. これらの意見を参考に, 文章を

表 2 読んだことがある作品に対する評価結果

番号	質問	1	2	3	4	5	平均
Q1	作品を的確に表したレビュー（感想）が表示されているといえるか	0	0	1	7	1	4.0
Q2	作品を的確に表したレビュー（作中の文章）が表示されているといえるか	0	0	0	5	4	4.4

表 3 読んだことがない作品に対する評価の結果

番号	質問	1	2	3	4	5	平均
Q1	レビュー（感想）のワードクラウドは本を選ぶ上で有効だといえるか	0	0	1	4	4	4.3
Q2	レビュー（作中の文章）のワードクラウドは本を選ぶ上で有効だといえるか	0	1	4	2	2	3.6

フェードインやフェードアウトなどのアニメーション表示にし、画像が移り変わる際の時間の間隔を伸ばすことで改善を行う。

5. まとめ

本論文では、評価の高いレビューと引用文をワードクラウドで表示を行う選書支援システムの開発を行った。しかし、現在、本システムはローカルネットワークでのみ動作が可能となっているため、ユーザによる文章の投稿、評価の部分が機能せず、書籍によってワードクラウドの質にばらつきが生じてしまう。そのため、青空文庫の書籍のみを対象とすることで、ワードクラウドの質を高め、多くのユーザに本システムを使ってもらった場面を想定して評価実験を行った。評価実験の結果、概ね高い評価をいただき、ワードクラウドの質が高い際の選書支援システムとしての有用性が明らかとなったが、文章の評価、投稿の機能がどれだけワードクラウドの質を左右させるかの検証は行うことができていないため、今後、本システムをオープンネットワークで公開することで、それらの問題を検証していきたい。また、現在は、ワードクラウドを生成するまでに1分ほどかかってしまうため、余計な処理を減らすなどして時間の削減を行っていきたい。

参考文献

- [1] 出版科学研究所, 日本の出版統計
<https://shuppankagaku.com/statistics/japan/>
- [2] 読書メーター, <https://bookmeter.com/>
- [3] ブックログ, <https://booklog.jp/>
- [4] 読書ログ, <http://www.dokusho-log.com/>
- [5] 藤田 佳恵, 鈴木 伸崇「小説データの本文分析による選書支援手法の提案」情報処理学会第 76 回全国大会
- [6] 村井 聡一, 牛尼 剛聡「興味喚起度 MAP を利用した電子書籍小説「立ち読み」インタフェースの実装と評価」DEIM2013
- [7] 毛利 勇摩, 寺田 実「極性辞書とワードクラウドを用いた商品レビューの可視化」DEIM2020
- [8] 上野 将, 市川 尚, 富澤 浩樹, 阿部 昭博「ワードクラウドを用いた授業中の掲示板書き込み内容を可視化するシステムの開発と評価」情報処理学会第 81 回全国大会 (2019)
- [9] ikegami-yukino (2019) mlask.py.
<https://github.com/ikegami-yukino/pymlask/blob/master/mlask/mlask.py>
- [10] R. Plutchik, “The nature of emotions,” American Scientist, Vol. 89, Iss. 4; pp. 344-350, 2001.

- [11] 夏目 漱石『ころも』青空文庫,
http://www.aozora.gr.jp/cards/000148/files/773_14560.html
- [12] 夏目 漱石『吾輩は猫である』青空文庫,
https://www.aozora.gr.jp/cards/000148/files/789_14547.html
- [13] 夏目 漱石『三四郎』青空文庫,
https://www.aozora.gr.jp/cards/000148/files/794_14946.html
- [14] 夏目 漱石『夢十夜』青空文庫,
https://www.aozora.gr.jp/cards/000148/files/799_14972.html
- [15] 太宰 治『斜陽』青空文庫,
https://www.aozora.gr.jp/cards/000035/files/1565_8559.html
- [16] 太宰 治『人間失格』青空文庫,
https://www.aozora.gr.jp/cards/000035/files/301_14912.html
- [17] 太宰 治『走れメロス』青空文庫,
https://www.aozora.gr.jp/cards/000035/files/1567_14913.html
- [18] 宮沢 賢治『銀河鉄道の夜』青空文庫,
https://www.aozora.gr.jp/cards/000081/files/456_15050.html
- [19] 坂口 安吾『墮落論』青空文庫,
https://www.aozora.gr.jp/cards/001095/files/42620_21407.html
- [20] 福沢 諭吉『学問のすすめ』青空文庫,
https://www.aozora.gr.jp/cards/000296/files/47061_29420.html

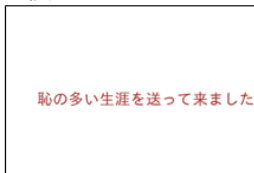
付 録

引用文を入力としたワードクラウド（人間失格）の構成画像を記載する。また、本論文で実験を行った各作品のワードクラウドは以下の url から参照可能である。

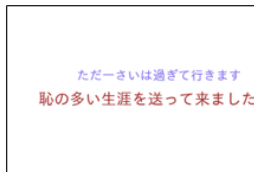
<https://drive.google.com/drive/folders/1MnXBTD8vANFW0vngJ-Qi9WtMUDyuo eos?usp=sharing>



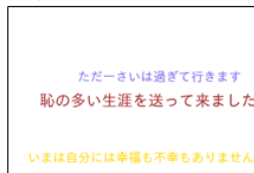
1枚目



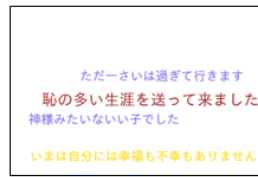
2枚目



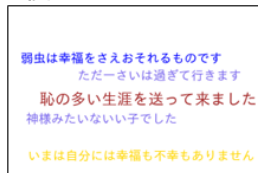
3枚目



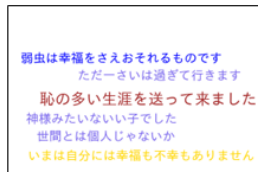
4枚目



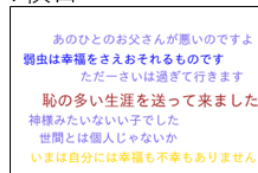
5枚目



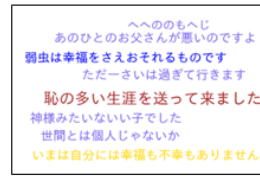
6枚目



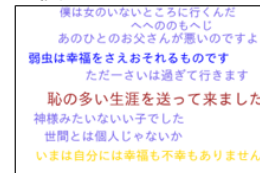
7枚目



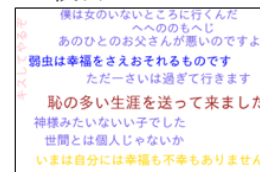
8枚目



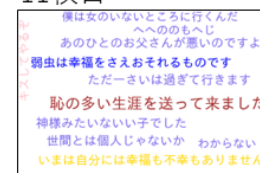
9枚目



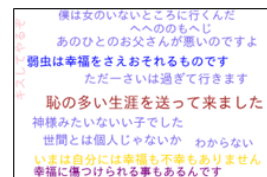
10枚目



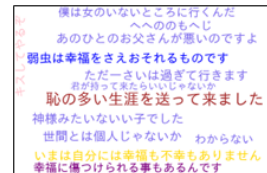
11枚目



12枚目



13枚目



14枚目