

電子書籍小説の選別における興味喚起度 MAP の有効性の評価

村井 聡^{1,a)} 牛尼 剛聡^{2,b)}

概要: 我々は、未読の電子書籍小説の選別支援を目的とした「立ち読み」インタフェースを提案している。本稿では、提案インタフェースに用いられている興味喚起度 MAP の有効性を評価する。実世界の書籍選別において、「立ち読み」は有効な手段とされているが、Web における電子書籍の「立ち読み」では、ユーザが閲覧可能なページが冒頭の数ページに制限されていることが多く、ユーザが興味をひかれる単語や文章を探すことが困難である。また、ユーザの検索クエリに応じて提示ページを変える試みもあるが、未読の小説に含まれる単語をクエリとして指定することは困難である。そこで、我々は、ユーザが電子小説の「立ち読み」を始める箇所を効率的に発見することを支援するために、「物語の進行における興味をひく度合いの推移」を可視化した興味喚起度 MAP を基盤とするインタフェースを提案してきた。被験者実験の結果、冒頭に近い層から順に選択される傾向があること、高い興味喚起度をとる層に配置された単語は、選択される度合いが高いことがわかった。

キーワード: 電子書籍提案 立ち読み、ブックレビュー、興味喚起

Abstract: We have proposed a browsing interface for supporting a user to decide whether he/she reads an e-book novel effectively. In this paper, we evaluate the effectiveness of attractiveness maps that are used in our interface. Book browsing is one of effective methods to select books in bookstores. However, on the web, book browsing is not effective. One of the reasons is viewable pages are limited. In some web pages, pages which are related to a query given by a user are displayed automatically. However, if users browse a novel, keyword retrieval based approaches are not effective because they would not like to know specified topics in the book previously. Our interface allows the users to select browsing sentences in a novel efficiently by an attractiveness map, which visualizes the transition of attractiveness in the novel according to its story flow. Our experiment showed users tend to select reading sentences from the left and to select terms which are in high peaks of attractiveness.

Keywords: E-books, browsing, book reviews, attractiveness.

1. はじめに

近年、書籍の電子化が進み、人は Web 上で書籍を探し、読むことができるようになった。しかし、候補となる書籍の増加にしたがって、読む書籍を見つけることが困難になり、書籍の効率的な選別を支援するシステムが重要になってきている。Web 上において、書籍を選別するタスクは、「候補を絞り込む段階」と「候補の価値を判別する段階」

の二つの段階に大別できる。ユーザが Web 上に存在する全ての書籍に目を通すことは困難であるため、まずはキーワードによる検索や推薦によって、候補を絞る作業を行うことが多い。しかし、一般的に、提示された候補の書籍を読む価値のある書籍と、即座に断定するのではなく、何らかの方法で価値を判断し、選別を行う。

選別方法の一つに、実際に書籍の本文を読む「立ち読み」がある。書籍等デジタル化推進事業の調査報告書 [11] によると、書店において、立ち読みを「よくする」と答えた人の数は、41.3%と最も多くなっており、書籍等の購入のきっかけとして、過半数の人 (50.8%) が「立ち読み・試し読み」を挙げている。その一方で、インターネット書店では立ち読み・試し読みを「全くしない」人の割合が 51.9%となっている。

¹ 九州大学大学院芸術工学府
Graduate School of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

² 九州大学大学院芸術工学研究院
Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

a) 2ds11094e@s.kyushu-u.ac.jp

b) ushiama@design.kyushu-u.ac.jp

その理由の一つとして、「立ち読み」時に閲覧可能なページが制限されているため、書籍の内容を判断することが困難となる場合があることがあげられる。現在、Web 上で提供されている電子書籍の「立ち読み」機構では、ユーザが閲覧可能な範囲が先頭ページから数ページに限定されていることが多い。そのため、書籍にユーザの興味をひく内容が書いてあったとしても、立ち読み時に閲覧できなかったために、読みたい書籍と判断できないこともある。この問題を解決するために、制限されたページ数内で、ユーザにとって有益な文章を多く提示できる「立ち読み」インタフェースが重要である。

電子書籍の「立ち読み」が利用できるサービスの一つである Google ブックス^{*1}[4] では、閲覧可能なページ数が制限されており、ユーザの入力クエリに応じて、提示するページを変更する。入力クエリによる「立ち読み」可能ページの提示は、専門書や実用書等を選別するような、ユーザが知りたい事柄があらかじめ把握している状況では有効的である。一方、未読小説の選別をする場合、小説本文に含まれる単語が未知であるために、ユーザは興味のあるキーワードの想起が困難であり、上記の手法は適切でない。

以上の背景より、我々は、「立ち読み」箇所を効率的に発見可能とすることを目的とした、電子書籍小説選別のための「立ち読み」用インタフェースを開発している。我々は、未読小説の選別においては、興味をひかれる箇所を発見できるかどうかを重要と考え、提案インタフェースでは「物語の進行」に基づいた「箇所がユーザの興味をひく度合い」の変化を可視化した「興味喚起度 MAP」を利用する。本論文では、興味喚起度 MAP の評価実験と、その結果を報告し、有効性について考察を行う。

本論文の構成は以下の通りである。第 2 章で関連研究を述べる。第 3 章で提案するインタフェースの概要について述べる。第 4 章で提案するインタフェースを実現するための手法を簡潔に述べる。第 5 章で提案するインタフェースで用いた興味喚起度 MAP の有効性を測る被験者実験と、その結果の考察について述べる。第 6 章で本稿のまとめについて述べる。

2. 関連研究

書籍選別に関する研究の中でも、書籍本文をブラウジングする行為に関する研究について幾つか述べる。松田 [5] は、図書館と書店における行動観察を行った結果、「ブラウジング」を「『出会わなければ必要な判断できない情報を含む、曖昧さを持つ情報欲求を満たすため、何らかの期待を抱きながら、利用できる感覚全てを用いて広範で多量な情報源から何らかの基準で必要なものを選び取る情報獲得の一手法である』」と定義している。本研究では、「立ち読

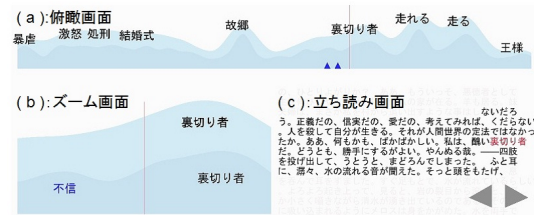


図 1 提案インタフェースによる『走れメロス』の表示例
Fig. 1 A snapshot of our book browsing interface on "Hahisre Merosu"

み」を『書籍を選別するために、書籍の本文を「ブラウジング」する行為』、効率的な「立ち読み」を『ユーザにとって、読みたいと思う書籍かどうかを、少ない時間、少ない文章量で判断すること』と定義する。また、松田は書籍を適当に開いて飛ばしながら見るという行為が、最も頻繁に観察された行為の一つであると指摘している。Annika[3] らは、学術的、調査目的に使われる図書館の利用者を観察し、彼らにインタビューを行った。Annika らは、書籍を選別する行為を、以下の 4 つに分類している。(1) カタログなどを介して、興味のある本を識別する。(2) 物理的に、棚から本を取り出す。(3) 使用可能な書籍の中から選択する。(4) 目的のコンテンツの書籍を読む。「立ち読み」は(3)に該当する。Annika らは手にとった書籍を、必要な書籍と判断するのに、目次、読書、フリッキングが役立っていたと報告している。小説の選別に限った研究には、Ooi[1] の研究があるが、「立ち読み」に関する記述はあまりない。また、電子書籍においては、Dana[2] は学術的な内容の書籍を対象にした、トランジションログの解析を行っている電子書籍の小説に対する「立ち読み」の有効性は未知数であるが、我々は、実世界で有効である「立ち読み」は、Web においても同様に有効だと考える。我々は Web における「立ち読み」においても、漠然とした興味がある段階では、小説を冒頭から提示するのではなく、全体を俯瞰できるようにし、興味を喚起する内容が目に触れることが重要だと考え、全体を俯瞰できる「立ち読み」インタフェースを提案する。

小説の可視化に関する研究として、山田 [10] の、シェイクスピアのストーリーの可視化があげられる。横軸に場面、縦軸に各場面の時間の流れをとり、ユーザが指定した、ポジティブな単語、ネガティブな単語の出現位置を基に、可視化を行う。この研究では、適切な単語を指定することで、ストーリーのテーマを把握することを可能としている。また、谷口 [6] は小説を深く鑑賞するために、三島由紀夫著『豊饒の海』の四部作それぞれに対する小説構造を、登場人物と、キーワードを利用した等高線グラフの文章地図による可視化を行っている。文書地図においては、表示する単語を小説本文に出現する頻度の高い単語の中から、ヒューリスティックを用いて抽出しているが、科学技

^{*1} <http://books.google.co.jp/>

術論文等と異なり、小説においては本文における出現頻度が高い単語が、本文の主題を適切に表しているとは限らない。Web 上にある全ての書籍に対して、あらかじめ人手で重要な単語と、その重要度を決定するのは現実的ではなく、また、ユーザが「立ち読み」を行う際に、未読書籍に対して、適切な単語を指定することも困難である。また、「立ち読み」においては、ストーリーの全体像以外にも、どこを読めばいいのか、同じ単語を含む場面が複数ある場合は、どちらを優先的に読めばいいのか等の情報も示す必要がある。我々は、Web 上にあるユーザのレビューを利用することにより、小説の主題に関わる単語から、読者の興味を喚起する可能性が高い単語を高精度に抽出し、それを基に小説の興味をひく度合いの推移を可視化する。

時間によって変わるデータをグラフとして可視化することで、全体像を効率的に把握し、探査する手法に関して、これまでいくつかの提案がなされている。Liu ら [7] は、電子メールのアーカイブを対象に、メール中に出現する単語頻度の時系列的な変化を可視化したインタフェースを提案している。このインタフェースでは、時間を横軸にとり、縦軸上にトピックの重要度が表現されている。本論文で提案するインタフェースでは、興味喚起度 MAP により、ユーザが小説の全体的な流れ俯瞰し、興味を持った箇所を閲覧可能とする。そして、ユーザが小説中の興味をひく内容を効率的に発見できるかを評価する。

3. 「立ち読み」インタフェース

本論文で提案する「立ち読み」インタフェースは、興味喚起度 MAP を基盤としている。興味喚起度 MAP は、「物語の進行」に伴って、「ユーザの興味を喚起する度合い」の推移をグラフで可視化し、グラフ上の特徴的な箇所にユーザの興味を喚起する単語を配置した図である。興味喚起度 MAP を利用することで、ユーザが興味をひく可能性が高い箇所を、「立ち読み」を開始する箇所として効率的に選択することを支援するインタフェースの開発を目指す。

図 1 は、太宰治著「走れメロス」を提案インタフェースのプロトタイプシステム上で閲覧している画面のスナップショットである。図 1 の (a) には興味喚起度 MAP を配置し、小説全体の流れを把握する俯瞰画面として用いる。興味喚起度 MAP の横軸は物語の進行を表し、縦軸はある箇所がユーザの興味を度合いの大きさを表している。層がピークとなる箇所には、その箇所において特徴的であり、興味を引くと予想される単語が配置される。(b) はズーム画面であり、ユーザが俯瞰画面において注目している部分(赤い線で選択している部分)を拡大して表示する。(c) は立ち読み画面であり、ユーザが注目する箇所の文章を表示する。俯瞰画面とズーム画面において、ユーザが単語を左クリックすると、立ち読み画面に、選択した単語を中心とした文章が表示され、ユーザは「立ち読み」を行うことが

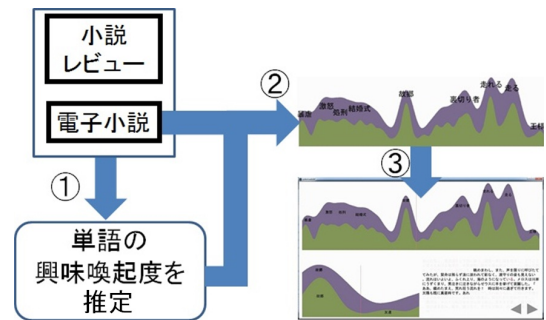


図 2 処理の概要

Fig. 2 An overview of processing

できる。また、層内で単語以外の箇所を左クリックした場合は、クリックした箇所にあたる形態素を中心とした文章が表示される。閲覧可能文章は、中心とする形態素から、前後 50 形態素となっている。例えば、図 1 は、単語「裏切り者」を左クリックした状況の様子を表している。また、画面右下の三角ボタンを押すことで、閲覧可能文章を読み進めたり、戻ったりすることが可能である。

4. 処理の流れ

本論文で提案するインタフェースにおける処理の流れを、図 2 に示す。最初に、小説本文に出現する単語の興味喚起度を、Web 上に存在する小説のレビューを使って決定する。対象とする小説のレビューに出現する頻度が高く、他の小説のレビューの出現頻度が低い単語に高い興味喚起度を与える。次に、興味喚起度 MAP を作成する。高い興味喚起度が密に出現する箇所が、高い興味喚起度をとるように可視化する。それにより、ユーザは興味を喚起する文章に出会う可能性が高い箇所を知ることができる。しかし、そこにどのような内容に関する文章が書かれているかは分からない。そこで、それぞれの層において、上に凸となる区間を見つけ、その区間において特徴的な単語を表示する。最後に、「立ち読み」に必要な機能を興味喚起度 MAP に付加したインタフェースを作成する。手法の詳細は、論文 [12] に示されている。

4.1 単語の興味喚起度

本手法では、Web 上のレビューを利用して、電子書籍小説に出現する単語の興味喚起度を推定する。レビューには、読者が書籍を読んだ感想が含まれており、読書中に興味を喚起された場面や人物に関する記述が含まれている。したがって、小説が読者に与えた印象という点から、小説の特徴を抽出できると考え、興味喚起度を推定する。

小説と、そのレビューの出現に基づいて、単語は以下の 4 種類のカテゴリに分類できる。

- 小説本文にもそのレビューにも出現する単語 (カテゴリ A)

- 小説本文には出現するが、そのレビューには出現しない単語 (カテゴリ B)
- 小説本文には出現しないが、そのレビューには出現する単語 (カテゴリ C)
- 小説本文にもそのレビューにも出現しない単語 (カテゴリ D)

ここで、カテゴリ A は小説の中で、レビューの興味を喚起した事項に関連する単語が含まれている可能性が高い。一方、カテゴリ B の単語は、小説本文には出現するが、ユーザの興味とは直接的に関係しない可能性が高いと考えられる。また、カテゴリ C の単語は、小説の情報、レビューの意見に関連する単語が多いと考えられる。また、カテゴリ D の単語は、小説の内容と関連する可能性は低いと考えられる。我々は、「レビューが興味を感じた対象は、他の読者の興味を引く可能性が高い」と仮定し、カテゴリ A に含まれる単語がユーザの興味を喚起する度合い（興味喚起度）を求める。なお、カテゴリ A 以外に含まれる単語の興味喚起度は 0 とする。また、登場人物名は、未読者の興味をひかないことが、我々の過去の被験者実験により示されているので、人物名の興味喚起度は 0 とする [12]。

ここで対象とする単語は、名詞（代名詞、接尾、数を除く）、動詞（自立動詞以外を除く）とする。ただし、他の語句の補助的な機能を持つ語句であることから、興味の強弱の判断材料として不向きであるため「する」、「ある」、「よる」、「いる」、「なる」、「いう」、「みる」、「できる」の 8 形態素 [9]、平仮名一文字、カタカナ一文字はストップワードとする。

4.1.1 興味喚起度の推定手法

対象とする小説とそのレビューの両方に含まれる単語が、ユーザの興味を喚起する可能性は同一ではない。本手法では、興味喚起度を小説における出現頻度ではなく、レビューにおける出現頻度に基づいて求める。具体的には、対象とする小説のレビューに出現する頻度が高い単語は、読者の興味を引いた度合いが高いとし、興味喚起度として高い値を与える。また、他の小説のレビューにも頻出する単語は、レビューに一般的に利用される概念であり、小説に関する可能性は低いと考える。

いま、書籍集合を B 、小説 b に対するレビューの集合を $R(b)$ とする。注目する小説を b としたとき、単語 t_i の興味喚起度 at を以下の式で定義する。

$$at(t_i, b) = \sum_{r \in R(b)} tf(t_i, r) \cdot \log \frac{|B|}{|\{b : r \in R(b), r \ni t_i\}|} \quad (1)$$

4.2 小説の特定箇所の興味喚起度とその変化

提案インタフェースでは、小説の興味喚起度の変化を提示することにより、ユーザが興味を持つ可能性が高い箇所

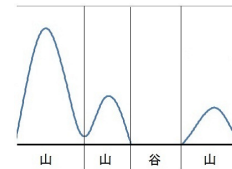


図 3 「山」と「谷」

Fig. 3 "Mountain" and "Valley"

を把握し、効率的に閲覧可能とする。本インタフェースでは、小説を構成する最小の意味的なまとまりは、小説中の形態素であると考え、「興味を喚起する単語を含む文脈は、興味を喚起する可能性が高い」と仮定し、高い興味喚起度をもつ単語が密に出現する箇所は、小説として興味喚起度が高いと考える。形式的には、小説に出現する形態素の出現位置を先頭から順番に i で表すとき、形態素出現位置 i を中心とした文脈の興味喚起度を、 $i-w \leq i, i \leq i+w$ の範囲に出現する単語の興味喚起度の重み付き和として求める。本手法では、形態素出現位置 i から離れた形態素の重要度を軽く扱うためにハニング窓関数 [8] を各単語の興味感輝度に乗じた。形式的には、形態素出現位置 i の文脈範囲 w の興味喚起度 $AT(i, w)$ を以下の式で定義する。

$$AT(i, w) = \sum_{c=i-w}^{i+w} at(i) \cdot \frac{1}{2} \left(1 + \cos 2\pi \frac{|c-i|}{2w} \right) \quad (2)$$

ここで、 $at(i)$ は形態素出現位置 i の単語の興味喚起度を示す。

本論文で提案するインタフェースでは、物語進行に伴う興味喚起度の変化を層として表現する。上位層は興味喚起度を計算する際の文脈範囲の値が大きく、下位層では興味喚起度の計算をする際の文脈範囲の値を小さくしている。つまり、文脈範囲の値は、ハニング窓の幅により決定し、ハニング窓の幅が多いときは大局的な興味喚起度の変化を表し、ハニング窓の幅が小さいときは局所的な興味喚起度の変化を表している。プロトタイプでは、立ち読み画面で提示する文章量の制約から、一番下層の窓幅を前後 100 語 ($w = 50$) に決定した。そして、窓幅を前後 200, 400 と 2 倍ごとに利用可能とした。

4.3 ユーザに提示する特徴的な単語の決定と配置

上記のように、本論文で提案するインタフェースでは、物語進行に伴う興味喚起度の変化を層として可視化することで、ユーザは興味を喚起する文章に出会う可能性が高い箇所を知ることができる。しかし、そこにどのような内容に関する文章が書かれているかは分からない。そこで、それぞれの層において、各ピークに特徴的な単語を表示する。本論文では、上に凸となる区間を「山」と呼び、興味喚起度が 0 となる区間を「谷」と呼ぶ。図 3 に「山」と「谷」の例を示す。

「山」に特徴的な単語は、その区間において出現する単

語の興味喚起度と、注目する「山」における出現頻度を考慮して決定する．具体的には、注目する層における全ての「山」からなる集合を M とし、全ての「谷」から構成される集合を V とするとき、ある「山」 m_j における単語 t_i の特徴量 un を以下の式で計算する．

$$un(t_i, m_j, w) = \left(\frac{at(t_i, w)}{\max_{t \in T}(at(t, w))} + \frac{tf(t_i, m_j)}{\sum_k tf(t_k, m_j)} \right) \cdot \log \frac{|M| + |V|}{|\{m : m \ni t_i, m \in M\}| + |\{v : v \ni t_i, v \in V\}|} \quad (3)$$

この特徴量 un は、小説の興味喚起度に基づいて、小説をいくつかの部分文書（「山」と「谷」）に分割し、それらの部分文書における単語の出現頻度に基づく局所的な重要度と、部分文書頻度の逆数に基づく大局的な重要度の積とした、単語の部分文書における重要度と解釈できる．

それぞれの「山」において、一番高い $un(t_i, m_j, w)$ をとる単語を、その「山」における特徴的な単語として表示する．しかし、同一の単語が何度も同じ層上に表示されるのは不適切であり、多様であることが望ましい．そこで、小説冒頭に近い「山」から順に、表示する単語を決定し、すでに表示された単語の興味喚起度 $at(t_i)$ を 0.1 倍し、同じ単語が複数回表示されないようにする．なお、その都度 $\max_{t \in T}(at(t, w))$ の値も更新している．

提案インタフェースにおいて、一つの層に表示させる単語の数は、最大 30 単語とし、高い層から優先的に表示させた．

5. 評価実験

5.1 実験手法

本論文で提案した興味喚起度 MAP の有効性を示すために、被験者実験に基づく評価を行った．実験では、興味喚起度 MAP 自体の有用性を見るために、プロトタイプシステムにおけるズーム画面を非表示にして実験を行った．また、機能の詳細は [12] に述べられているが、単語検索機能、層の絞り込み機能を使用不可とした．実験に用いる小説のデータは、青空文庫^{*2}で公開されている小説の内の、40 冊分を使用した．レビューは web 本棚サービスである^{*3}に投稿されているものを使用した．形態素解析には、オープンソースの形態素解析エンジン MeCab^{*4} を用いた．

被験者実験では、「立ち読み」と「実読書」の 2 種類の状況を想定する．それぞれの状況において、被験者には興味をひかれた箇所を指定してもらった．我々は、「実読書」中に発見された興味をひかれたページを正解データとし、「立ち読み」中に発見された興味をひかれた文章と比較することで、興味喚起度 MAP の有効性を評価する．

興味喚起度 MAP の評価のために、比較対象として 2 つ

^{*2} <http://www.aozora.gr.jp/>

^{*3} Booklog<http://booklog.jp/>

^{*4} <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

のマップを用意し、以下の 3 種類のマップを利用して実験を行った．

- (a) ブロック型 MAP: 横軸が物語の進行を表し、単語を配置したマップである．このマップでは、興味喚起度の大きさが可視化されていないため、キーワードの提示による効果が反映される．ブロック内では、単語が重ならないよう、斜めに配置されているが、その高さには意味は無い．
- (b) 興味喚起度 MAP: 横軸が物語の進行を表し、縦軸に興味喚起度を表した興味喚起度 MAP（提案手法）．
- (c) ランダム MAP: 横軸上に物語の進行を表し、縦軸が興味喚起度を表すマップである．興味喚起度 MAP との違いは、単語の興味喚起度の順序を乱数によって入れ替え、興味喚起度の値が本文の内容を反映していない点である．これは、興味喚起度マップの結果と比較することにより、興味喚起度の提示が与える影響を知ることができる．

夏目漱石著「こころ」を例として、上記の 3 種類のマップの例を図 4 に示す．我々は、「興味喚起度 MAP は、小説の興味をひく度合いの推移を正しく示しているため、特徴的な単語を選ぶ助けとなり、より多くの興味をひく場面を発見することができる」と仮説を立てた．なお、ブロック型 MAP とランダム MAP 内に表示する単語は、興味喚起度 MAP におけるものに統一した．また、単語の横軸の座標は、ブロック型 MAP では、各ブロックの中心座標、ランダム MAP では、興味喚起度 MAP と同じ座標を用いた．

5.1.1 「立ち読み」による興味箇所の選択

まず、被験者には、小説 4 題（太宰治著「斜陽」（書籍本文:199 ページ）、ポー著「モルグ街の殺人事件」（51 ページ）、太宰治著「人間失格」（162 ページ）、夢野久作著「ドグラ・マグラ」（637 ページ））を、指定した MAP を用いたインタフェース上で「立ち読み」を行なってもらう．「立ち読み」時間は 2 分間とし、できるだけ多くの興味をひく箇所を発見してもらった．興味をひく箇所を発見した際には、キーボードのキーを押してもらった．実験は、全て同じ PC 上で行った．また、実験で利用した層の窓幅は、斜陽、モルグ街の殺人事件、人間失格では前後 1600 語と前後 3200 語とした．ドグラ・マグラにおいては、ページ数が他の三冊と比較して多く、層の推移をより緩やかにするために、窓幅を前後 3200 語と前後 6400 語とした．

5.1.2 「実読書」による興味箇所の選択

本研究では、実際の読書時に興味をひかれたページを正解データとして用いた．5.1.1 の「立ち読み」実験の後、被験者には、各小説を実際に紙媒体で読書してもらい、「興味をひかれた内容が書かれているページ」に付箋を貼ってもらい、それらを正解データとした．興味をひかれた範囲を厳密にとるのは、自然な読書の妨げとなるため、興味をひかれた箇所の単位は、ページとした．また、その場限りで

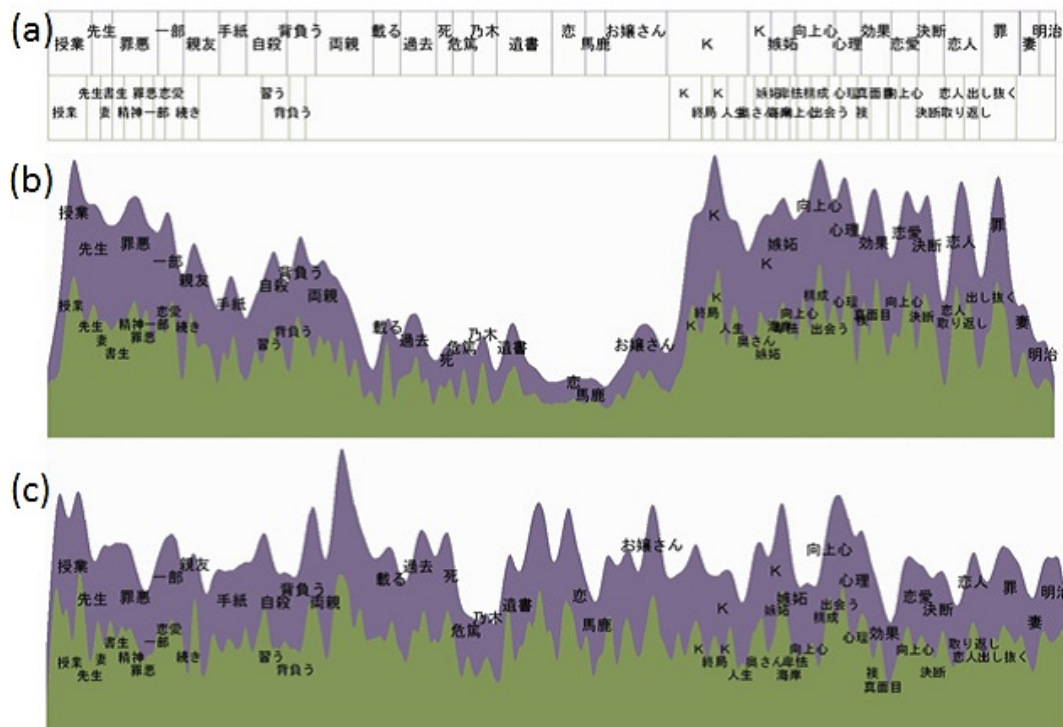


図 4 評価実験に用いるマップによる「こころ」の表示例
Fig. 4 Example of the maps on "Kokoro"

の興味をとるために、以下の 3 点を守ってもらった。

- (1) 飛ばし読みをせず、冒頭から順に読み進める。
- (2) 前のページに戻って、付箋を貼ったり、はがしたりはしない。
- (3) 付箋の付与を迷い決められない場合は、付箋を半分にちぎって貼る。

5.1.3 正解の判定

本実験における正解の判定について述べる。上記で説明した「立ち読み」で選択された本文中の箇所と、「実読書」で選択された正解ページは、そのままでは比較することが出来ない。そこで、「立ち読み」で選択された箇所を、書籍上でそれを含むページへと変換する。「立ち読み」においては、被験者は、任意の単語、もしくは、層の一部、を選択することにより、「立ち読み」を開始する箇所を選択する。そして、多くの場合において、しばらく読み進め、興味をひかれる箇所を発見した、もしくは、興味をひかれなないと判断した際に、他の単語、層の一部を選択し直す。「立ち読み」箇所を選択してから、興味をひかれた箇所を発見、もしくは、他の単語、層の一部を選択するまでを、一つのチャンクとする。各チャンクで一番最後に表示された文章が含まれるページを正解データとの比較に用いる。なお、最後に表示された文章が複数ページにまたがる場合は、両ページを比較に用いる。図 5 に比較に用いるページの決定の例を示す。記録された興味をひかれた箇所と、「立ち読み」を終了した箇所は、それぞれ興味があった箇所、興味

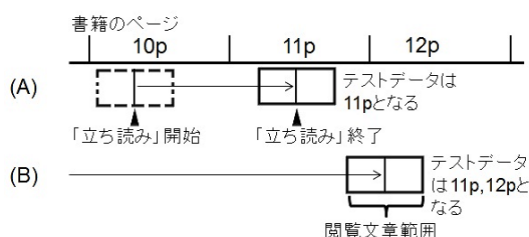


図 5 「立ち読み」箇所のページへの変換
Fig. 5 Convert browsed sentences into pages

がない箇所、ととらえることが出来るが、「立ち読み」を進めた箇所については、興味があったとも、なかったともとらえることができるため、比較には含めないものとした。

5.2 実験結果

9 名の被験者を対象に「立ち読み」の実験を行った。対象とした書籍の題名と、対象としたマップを利用した被験者の人数を表 1 に示す。また、「実読書」を行った被験者の人数、付箋の数を表 2 に示す。

被験者 1 以外の被験者は、付箋を半分にちぎって貼るという行為をほとんどしなかった。そこで、ちぎらずに貼られた付箋のみを正解データとした。各書籍の適合率と再現率を表 3 に、各マップごとの適合率の平均と再現率の平均を表 4 に示す。

マップの種類による、適合率と再現率に関する有意な違いは見られなかった。

被験者の「立ち読み」行動観察から、「立ち読み」開始箇所

表 1 「立ち読み」の種類と被験者数

Table 1 Types of stand reading and numbers of subjects.

書籍タイトル	ブロック型	興味喚起度	ランダム	計
斜陽	3	3	3	9
モルグ街の殺人事件	2	3	2	7
人間失格	3	4	2	9
ドグラ・マグラ	2	2	4	8
合計	10	12	11	33

表 2 「実読書」を行った被験者数

Table 2 Numbers of subjects for actual reading

被験者 id	書籍	付箋 (数)	付箋半 (数)
1	斜陽	13	33
2	斜陽	13	0
3	斜陽	24	0
11	斜陽	15	0
1	モルグ街	6	0
2	モルグ街	8	0
3	モルグ街	6	1
5	モルグ街	7	1
9	モルグ街	2	0
10	モルグ街	19	0
1	人間失格	12	23
1	ドグラ・	20	36

表 3 適合率と再現率

Table 3 Recall and precision.

書籍	MAP	適合率	再現率	比較数	正解数
斜陽	block	0.60	0.12	10	65
斜陽	興味喚起	0.64	0.27	22	65
斜陽	ランダム	0.70	0.13	10	65
モルグ街	block	0.21	0.1	7	48
モルグ街	興味喚起	0.31	0.33	16	48
モルグ街	ランダム	0.35	0.47	20	48
人間失格	block	0.46	0.17	23	12
人間失格	興味喚起	0.46	0.17	25	12
人間失格	ランダム	0.48	0.08	28	12
ドグラ・	block	0	0	27	20
ドグラ・	興味喚起	0	0	7	20
ドグラ・	ランダム	0	0	13	20

表 4 マップ毎の再現率の平均と適合率の平均

Table 4 Each map's average recall and average precision.

MAP	適合率	再現率
ブロック型	0.32	0.1
興味喚起度	0.35	0.19
ランダム	0.38	0.17
平均	0.35	0.11

所を選択する順序として、層の高い順ではなく、冒頭に近い順から選ぶ被験者が多い傾向があった。興味喚起度 MAP において高い値をとる層、ランダム MAP において高い値をとる層に配置された単語の幾つかは、他の MAP に配置

表 5 「立ち読み」開始箇所の内訳

Table 5 The number of selected objects on maps

選択対象	興味をひかれた数	興味をひかなかった数	合計
単語	113	182	295
層	29	73	102
合計	142	255	397

「実読書」において興味を喚起したページ	
興味をひいた	
興味をひかなかった	
「立ち読み」において興味を喚起したページ	興味をひいた
	興味をひかなかった
	興味をひかなかった
「立ち読み」において興味を喚起したページ	興味をひいた
	興味をひかなかった
	興味をひかなかった
「立ち読み」において興味を喚起したページ	興味をひいた
	興味をひかなかった
	興味をひかなかった

図 6 「立ち読み」時と、実読書時の、興味をひかれた箇所の分類

Fig. 6 Classification based on reading pages and browsing pages

された同じ単語と比較して、選択回数が多かった。また、興味喚起度 MAP とランダム MAP の両者において低い値をとる箇所に配置され、一度も選択されなかった単語が、ブロック型 MAP において複数回選択されたケースもあった。以上のことから、層により、ユーザの読むべき箇所を誘導することはある程度可能であることがわかった。

「立ち読み」を開始する箇所の選択方法の内訳を表 5 に示す。任意の単語選択によるものが 295 箇所あり、層の一部を選択したものが、102 箇所あった。このことから、被験者がマップ上の単語の選択により立ち読み箇所を決定する傾向が強いことがわかる。また、興味をひかれた箇所につながった割合は、単語選択が 0.38、層選択が 0.28 であった。

「立ち読み」開始箇所を選択して、表示された文章に即座に興味をひかれたチャンク数は 3 件、また、即座に別の単語、層の一部を選択したチャンク数は 8 件であった。多くのチャンクにおいて、読み進める行為を行っており、「立ち読み」画面を改善することで、より使いやすいインタフェースになると考えられる。

5.3 考察

実読書時に興味をひかれたページは、「立ち読み」時に興味をひかれた箇所を含むページとの関係から、図 6 に示す 6 種類のカテゴリに分類できる。カテゴリ A は読書時、「立ち読み」時ともに、被験者の興味をひいたページであり、「立ち読み」時に読む価値がある箇所である。カテゴリ

Bは、「立ち読み」では、読者の興味を引いたものの、読書時には興味をひかなかったページである。実読書の際には興味をひかれなかったことから、物語の本筋に関係ない、もしくは、台詞や表現などが印象的に面白かった箇所だと考えられる、これらは、適合率を下げる要因になるが、効率的な立ち読みを支援する際には重要な影響を与えないと考えられる。

カテゴリ C_r は、「立ち読み」において被験者が興味をひかれなかったが、読書時に被験者の興味をひいたページである。このページは、物語の筋を知らなければ興味をひかれない箇所である可能性がある。被験者毎に結果を見ると、このカテゴリに含まれるページ数は最大8ページであり、平均1.68ページとそれほど多くはない。しかし、適切な順で「立ち読み」をすることが出来れば、「立ち読み」において興味をひかせることが出来る可能性があるため、今後の課題とする。

カテゴリ C_u は、読書した時に興味をひいたページであるにも関わらず、「立ち読み」では閲覧ページとして選択されなかったページである。興味をひく可能性がある箇所が、「立ち読み」で目につかない、というのが、本論文において指摘している問題であり、解決する必要がある高い箇所である。提案手法では、レビューでの出現頻度により決定した、興味喚起度を基にして、優先的に閲覧すべき箇所を可視化している。しかし、カテゴリ C_u に含まれる箇所には、レビューには出現しなかったり、出現したとしても出現頻度が低い単語が多く含まれていることもある。別の手法で、このカテゴリに含まれるページを減らしていくことも今後の課題である。

カテゴリ D_r は、「立ち読み」において読者が読んだ結果、興味を惹かれなかった文章であり、読書において、被験者の興味をひかなかったページである。このページは、「立ち読み」で読む優先順位が低いページであり、「立ち読み」において読む必要がなかったページである。これらは、適合率を下げる要因であり、このカテゴリに含まれるページを減らすことが重要である。

カテゴリ D_u は、「立ち読み」において読まれなかったページ、かつ、読書において、被験者の興味をひかなかったページである。このページは、「立ち読み」で読む優先順位が低いページであり、「立ち読み」において読まなくてよかったページである。

今後の研究では、上記の分類に基づき、興味をひかれた文章、ひかれなかった文章の特徴を考察し、インタフェース改善していく予定である。

6. まとめ

本稿では、電子書籍小説を対象とした「立ち読み」インタフェースにおける、興味喚起度 MAP の評価を行った。被験者実験の結果、提案するマップと、比較に用意した

マップとの間に、再現率、適合率における違いは見られなかった。しかしながら、被験者が、冒頭に近い箇所から順に「立ち読み」を進める傾向があること、高い層に配置された単語を選択する傾向があることがわかった。また、被験者は、表示された文章が含まれる場面に、興味をひかれるかどうかを、即座に判断するのではなく、本文を読み進めてから判断する傾向があることがわかった。今後は、興味をひかれる可能性が高い箇所をより正確に示せるように手法を改善するとともに、インタフェースを層を中心としたものから、単語、書籍本文を中心としたものへと改善していく予定である。

謝辞 本研究は JSPS 科研費 24500119 の助成を受けたものです。

参考文献

- [1] Ooi, Kamy.:How adult fiction readers select fiction books in public libraries: a study of information-seeking in context, Masters, Victoria University of Wellington, New Zealand(2008)
- [2] McKay, Dana.: A jump to the left (and then a step to the right): reading practices within academic ebooks, *OzCHI '11*, pp. 202-210 (2011).
- [3] Hinze, Annika., McKay, Dana., Vanderschantz, Nicholas., Timpany, Claire and Cunningham, Sally Jo.: Book selection behavior in the physical library: implications for ebook collections, *JCDL '12*, pp. 305-314 (2012).
- [4] 村上 憲郎:Google の切り開く情報の世界 プロフェッショナルの仕事とは、情報管理, Vol 53, No 12, pp. 651-664 (2011).
- [5] 松田 千春:情報探索におけるブラウジング行動: 図書館と書店における行動観察を基にして, 三田図書館・情報学会, No 49, pp. 1-31 (2003).
- [6] 谷口 敏夫:『天人五衰』の小説構造可視化: 三島由紀夫『豊饒の海』の絵解き, 京都光華女子大学研究紀要, Vol. 42, pp. A35-A57 (2004).
- [7] Liu, Shixia., Zhou, Michelle X., Pan, Shimei., Qian, Weihong., Cai, Weijia., Lian, Xiaoxiao: Interactive, topic-based visual text summarization and analysis, *CIKM '09*, pp. 543-552 (2009).
- [8] 黒橋 禎夫, 白木 伸征, 長尾 眞: 出現密度分布を用いた語の重要説明箇所の特定, 情報処理学会論文誌, Vol 38, No 4, pp. 845-854 (1997).
- [9] 沢井 康孝, 山本 和英: 文書に対する大衆の興味の強さの推定提案人工知能学会全国大会論文集 (CD-ROM) (2007)
- [10] CPS: 山田 美幸: プラネット・シェイクスピア: 文学作品におけるストーリーの可視化 (online), 入手先 (https://www.cps-jp.org/modules/mosir/player.php?v=20110223_yamada) (2011.02.23).
- [11] 日本出版インフラセンター: 電子出版と紙の出版物のシナジーによる書店活性化事業 (online), 入手先 (http://www.meti.go.jp/policy/mono_info_service/contents/downloadfiles/120307-1.pdf) (2012.02.19).
- [12] 村井 聡一提案牛尼 剛聡: 電子書籍小説の効率的な選別のための興味喚起度に基づく「立ち読み」インタフェース提案 WebDB Forum (2011).