Vol.2017-IFAT-125 No.8
Vol.2017-DC-104 No.8
2017/3/10

EPUB を用いた読書活動支援システムの開発

外川大悟[†1]　大場みち子[†2]

概要：近年, インターネット上で読書に関する情報を多人数で共有するソーシャルリーディングサービスが注目を集めている. ソーシャルリーディングによって共有された感想や書評などの情報から, ユーザに適した情報提供をどう行うかが課題となっている. 本研究では, 書籍情報を効率的に探すための情報提示手法の提案を目的とする. この目的を達成するため, 書籍内のアノテーションから書籍に対する感情やキーワードを抽出し, これらのビジュアル化を行う. ビジュアル化されたこれらの情報を共有することで, 情報提示を効率化させユーザの読書活動の活発化を図る. 実験により提案手法の有効性を確認する.

キーワード：ソーシャルリーディング, アノテーション, 読書活動, ビジュアル化, 感情

Development of the Reading Activity Support System Using EPUB

DAIGO TOGAWA[†1] MICHIKO OBA[†2]

Abstract: Social reading services are receiving a lot of attention recently. Social reading service is a system that shares information on reading on the Internet with many people. The challenge for social reading systems is how to provide users with adequate information such as shared impressions and book reviews. The purpose of this study is to activate reading activities of users by efficiently conveying useful information for them. In order to achieve this purpose, we extract emotions and keywords for books from annotations in books and make visualization of them. Users share these visualized information with each other. Then, system presentations of information more efficient and activate the reading activity of the users. Finally, we confirm the effectiveness of the proposed method by experiment.

Keywords: Social reading services, Annotations, Reading Activity, Visualize, Emotion

1. はじめに

最近ではソーシャルリーディングサービスという, 感想や書評など書籍に関する情報を多数のユーザで共有するサービスが注目されている[1]. Kindle[2]や読書メーター[3]などを利用することで多くの書籍の情報を集め, 読みたい書籍を見つけることでユーザの読書活動が活発化される.

また, 最近ではソーシャルリーディングシステムで共有される情報の種類が増えている. その中で注目されているのがアノテーションである. アノテーションとは文章中に関連する情報を注釈として付与することである. ソーシャルリーディングシステムでは, 文章の一部を反転表示や背景色を変えて強調表示を行うハイライト, 文章中に感想や注釈を書き込むメモといった情報のことを指す. こういったアノテーション情報はユーザが重要だと考えた部分につけられることが多く, アノテーション情報を共有することによって, ユーザは書籍の重要な要素を共有することが可能となる.

しかし, 全てのユーザが自分の探したい書籍情報のキーワードを理解しているわけではない. 中には, 書籍を探す際に「面白い」や「楽しい」などイメージで検索を行いたいユーザが存在する. こういったユーザは目的の書籍をキーワードで探すことができず, 求めている情報を得るのが難しい. 現在のソーシャルリーディングシステムで, 「面白い」や「楽しい」といったイメージで書籍に関する情報を検索すると, 対象となる情報が多すぎるためユーザは目的の情報を探し出すことができない. つまり, 現在のソーシャルリーディングよる情報共有には, 有効な情報共有がなされていないユーザが存在するという問題がある.

そこで, 本研究では書籍情報を効率的に探すための情報提示手法の提案を目的とする. この目的を達成するためにユーザに対して有効な書籍情報の提示を行うことで, ユーザの読書活動を支援するシステムを提案する.

本稿では, 最初に先行研究とその問題点, 本研究の目的と課題について述べる. 次に, 課題の解決アプローチについて述べる. その後, 今回行った本研究の有用性の見通しを確認する予備実験方法について述べる. 最後に, 実験結果とその考察, 今後の研究方針について述べる.

2. 関連研究

片岡らは世界的に普及している電子書籍フォーマット EPUB (electronic publication)[4]を用いて, 読書に関する情報を多人数で共有するソーシャルリーディングシステムを開発している[5]. また, 開発したシステムを用いて実験を行

†1 公立はこだて未来大学 大学院
Future University Hakodate Graduate School
†2 公立はこだて未来大学
Future University Hakodate

い，収集した情報を基にユーザ分析を行なっている．ユーザのブックマークやアノテーションを分析することにより，書籍内でのユーザの興味・関心を知ることができる．また，この研究では EPUBCFI (electronic publication Canonical Fragment Identifier)[6]と呼ばれる部分文書識別子を利用している．これを用いることで，EPUB 電子書籍内の任意の一点を唯一に識別することが可能となる．この技術によって EPUB 内のハイライト，メモ，注釈などの場所を参照し，これらのアノテーション情報をユーザ同士で共有することにより，この研究ではソーシャルリーディングを実現している．アノテーション情報によってソーシャルリーディングで共有できる情報の種類を増やし，その情報の利用法が提案されている．

王らは研究室向けに特化した，学術論文に対するソーシャルリーディングを実現できる文献評価システムの構築をしている[7]．この文献評価システムは，ソーシャルリーディングでよく用いられるアノテーション機能を，PDF ファイル対して実装している．また，このシステムはウェブ技術で実装しているため，ユーザはウェブブラウザ上で文献を参照・注釈の追加が可能となる．この研究では，アノテーションの共有手法と，その実装方法が提案されている．

しかし，これらの研究によって共有される情報は文章であることが多く，ユーザは情報を探す際，共有された文章を読む必要があり，自身が必要としている情報を探すのに時間がかかってしまう．つまり，ユーザに対して効率的な情報提示手法が求められている．そこで，本研究では書籍情報を効率的に探すための情報提示手法の提案を目的とする．この目的を達成するための課題を以下に示す．

(1) 書籍を探すために書籍のどのような情報を利用するか
(2) 収集した書籍の情報をどう効率的に表示するか

3. 解決アプローチ

3.1 概要

本研究では 2 章で述べた課題を解決するため二つの解決アプローチを提案する．まず，2 章の課題(1)の解決アプローチとして，アノテーション情報の利用を提案する．前述の通り，アノテーション情報は書籍内で読者が重要だと考えた部分につけられることが多いため，この情報を利用することによって多くのユーザが重要だと考えた本の情報を提示することができる．また，ユーザは本を読むことなく本の重要な要素を知ることができる．

課題(2)の解決アプローチとして，アノテーション情報のビジュアル化を提案する．アノテーション情報は文章であり，ユーザが必要な情報を検索・取得するのに時間がかかってしまう．そこで，本研究ではアノテーションから収集した情報を，グラフやタグクラウドといった形にビジュアル化する．ビジュアル化によってユーザは文章を読む必要がなくなり，書籍の情報取得・検索にかかる時間が短縮される．このビジュアル化によって，収集した書籍の情報を効率良く提示することが可能となる．

3.2 ビジュアル化手法

本研究のビジュアル化手法を 2 種類提案する．これら二つのビジュアル化手法を用い 2 章の課題(2)を解決する．

3.2.1 タグクラウドによるキーワード表示

本研究では，タグクラウドを用いたキーワード表示を提案する．タグクラウドとは，図 1 でに示すようなウェブサイト上などで使われるタグの視覚的記述のことである[8]．タグの頻出度によってフォントや表示サイズ，色を変更することでそれぞれのタグの重要度を表現することができる．本研究ではこのタグクラウドを利用し，書籍内のキーワードの重要度がわかるように表示する．

図 1: タグクラウドの例
Figure1: Example of Tag Cloud

3.2.2 書籍内感情のグラフ化

本研究では，ビジュアル化手法として書籍内感情のグラフ化を提案する．この手法では，アノテーションが付けられた文章やコメントを抽出し，抽出された情報内に，どういった感情が，どの程度含まれているのかを分析する．その結果を図 2 のようにそれぞれの感情の割合を色分けした円グラフで表示する．主に感情の割合を表示する感情グラフによって，読者は書籍を読むことなく，対象の書籍にどういった感情が含まれているのか効率良く理解することができる．

Vol.2017-IFAT-125 No.8
Vol.2017-DC-104 No.8
2017/3/10

4. 提案システム

4.1 概要

本研究では以上の解決アプローチが可能となるシステムの構築を目指す．構築するシステムの利用手順を図 3 に示す．

(1) ユーザは，自身が読みたいと考えている書籍をイメージしつつシステムの利用を開始する．
(2) ユーザの書籍内に付けられたハイライトが付けられた文章や，文章中に付けられたメモといった，アノテーション情報を抽出する．
(3) 抽出された情報から文章中のキーワードや，文章中に含まれる感情を抽出しそれらをビジュアル化する．
(4) ユーザはビジュアル化された情報を見ることによって，その書籍ではどういった単語がキーワードとなっているのか，また書籍内にどういった感情が含まれているのか，などを直感的に理解することができる．
(5) 共有されたこれらの情報を見ることによって，ユーザは書籍を読まずに書籍の情報を効率良く獲得し，従来の手法よりも早く目的の書籍情報を探すことができる．

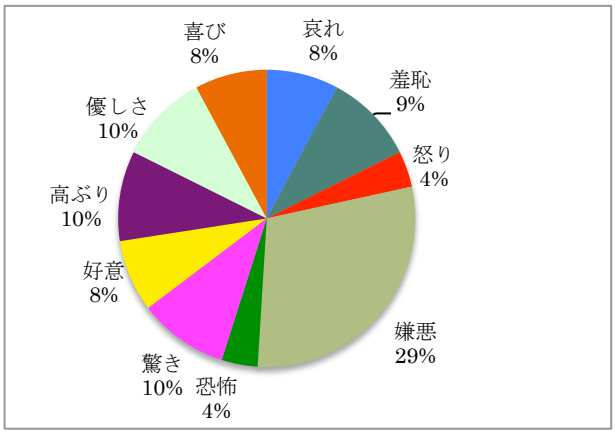

図 2: ML-Ask を用いた書籍内感情グラフ例
Figure2: Example of Emotion Graph in Book Using ML-Ask

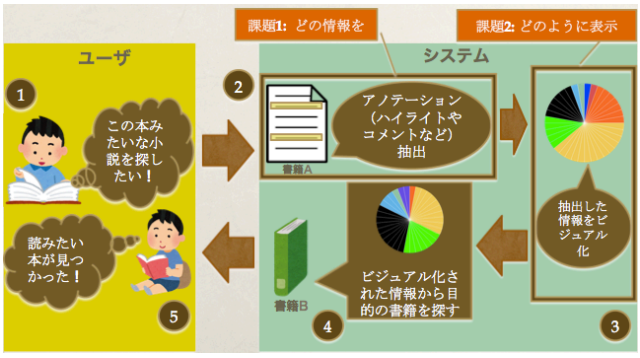

図 3: システム利用の流れ
Figure3: Flow of system use

4.2 キーワードの抽出方法

本研究では，書籍中に付けられたアノテーション情報を抽出し，分析するために，抽出した情報に対して日本語形態素解析エンジンである MeCab[9]を用いて形態素解析を行う．形態素解析によって文章を単語単位まで分割し，その結果に対して TF-IDF を用いてキーワードの抽出を行う．TF-IDF とは主に情報検索や文章要約で利用される，該当文章内に出現する語の頻度の情報をもとに重要度を決定する単語に関する重みの一種である[10]．抽出したキーワードの重要度の算出結果を基準にタグクラウドで表示するサイズを決める．重要度が大きいほど表示サイズは大きく，重要度が低いほど表示サイズは小さい．このビジュアル化手法によって，ユーザは書籍を読まずに書籍内のキーワードを理解することができる．

4.3 文章内感情の分析方法

本研究では，抽出されたアノテーション情報に対して ML-Ask という感情解析プログラムを用いる．ML-Ask とは文章中に存在する感情を表す単語を検出し，その単語を哀れ，羞恥，怒り，嫌悪，恐怖，驚き，好意，高ぶり，優しさ，喜び 10 種類の感情に分類し，それぞれの感情を表す単語の個数を出力するプログラムである[11][12]．このプログラムを用いることで対象の文章にどういった感情が，どの程度含まれているのかがわかる．この結果をもとに，図 2 のような書籍内感情のグラフを作成することが可能となる．

5. 予備実験

5.1 概要

今回提案した手法の有用性の見通しを確認するために予備実験を行った．予備実験の目的はアノテーションを利用した情報と，従来の書評とのビジュアル化に対する効果の違いを検証することである．今回の実験では，利用するアノテーション情報として，ハイライトのみを対象とする．文章中に付けられたハイライトを利用してビジュアル化した情報と，Amazon の評価レビューを利用してビジュアル化した情報を比較する．ビジュアル化したそれぞれの情報の特徴や相違点を確認し，今回の提案手法の有用性を検証する．今回の実験では「芥川龍之介」の殺人事件をテーマとした短編小説である『藪の中』を対象とした．対象としたハイライト数は 98，アマゾンの評価レビュー数は 38 である．また，今回は公立はこだて未来大学内の教員・学生が，ハイライト付けをした．

5.2 実験結果

5.2.1 タグクラウドによるキーワード表示

アノテーションから作成したタグクラウドを図 4 に，Amazon レビューから作成したタグクラウドを図 5 に示す．アノテーションから作成したタグクラウドでは「妻」，「男」，

「小刀」，「盗人」など書籍中に含まれるキーワードが多く表示された．Amazon レビューから作成したタグクラウドでは「映画」，「羅生門」，「原作」，「芥川」など作品に関連する情報が多く表示された．他にも「面白い」，「素晴らしい」などこの書籍を読んだ読者の主観的な感情を抽出することができた．

図 4: タグクラウド実験結果（アノテーション利用）
Figure4: Experimental result of tag cloud (using annotation)

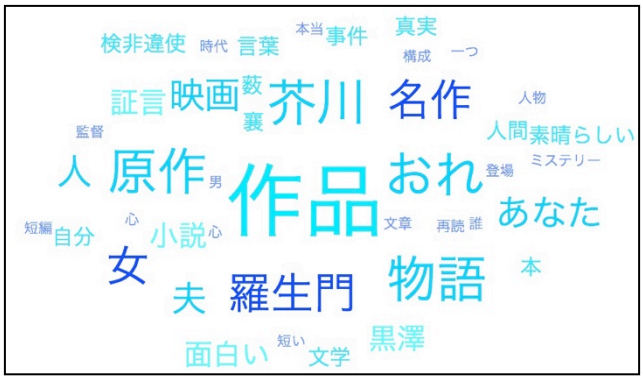

図 5: タグクラウド実験結果（Amazon 評価レビュー利用）
Figure5: Experimental result of tag cloud (using Amazon review)

5.2.2 ML-Ask を用いた感情グラフ

アノテーションから作成した感情グラフを図 6 に，Amazon レビューから作成した感情グラフを図 7 に示す．アノテーションから作成した感情グラフでは「羞恥」，「嫌悪」など書籍内の登場人物の感情が多く表示された．Amazon レビューから作成した感情グラフでは「好意」，「喜び」など読者の書籍を読んだ際の感情が多く表示された．

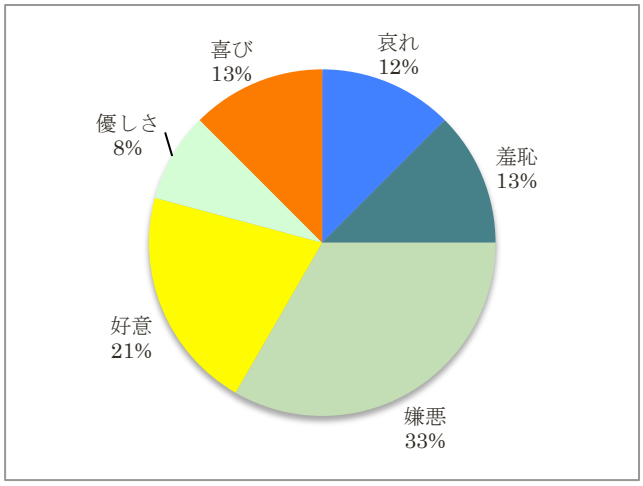

図 6: 感情グラフ実験結果（アノテーション利用）
Figure6: Experimental result of emotion graph (using Annotation)

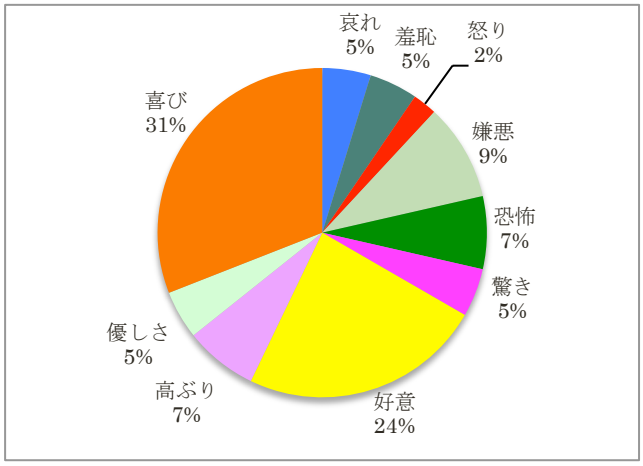

図 7: 感情グラフ実験結果（Amazon 評価レビュー利用）
Figure7: Experimental result of emotion graph (using Amazon review)

6. 考察

今回の予備実験によって情報を抽出する媒体と表示法の違いで得られる情報が異なるということが判明した．アノテーションから抽出した情報をビジュアル化したタグクラウドと感情グラフは書籍自体の情報を多く含んでいた．Amazon レビューから抽出した情報をビジュアル化したタグクラウドと感情グラフからは，書籍を読んだ人の感想や書籍に関連する情報が多く含まれていた．これらの結果から，本研究で提案した手法は従来の手法とは違う情報を提供することが可能である．

しかし，今回の予備実験で作成したタグクラウドはユーザにとって必要のない情報が多く表示されていた．この結果から，タグクラウドで表示する際に必要のない情報を排除する必要がある．ML-Ask を用いた感情分析を行う際，感

情ごとに区別することができる単語の数が少なく，文章内の感情を正確に分析することができない．今後は分類することができる単語の種類を増やすことなどで，ML-Ask による感情分類の正確性向上を図る必要がある．

7. まとめ

本研究では書籍情報を効率的に探すための情報提示手法の提案を目的とし，ユーザに対して有効な書籍情報の提示を行うことで，ユーザの読書活動を支援するシステムの開発を目指した．これらを達成するため，ハイライトされた文章や，添付されたメモなどのアノテーション情報をビジュアル化し，それらをユーザ同士で共有する手法を提案した．また，ビジュアル化の手法としてタグクラウドによるキーワード表示と，書籍内感情のグラフ化の 2 種類を提案した．今回提案した手法の有用性を検証するために，文章中に付けられたハイライトを利用してビジュアル化した情報と，Amazon の評価レビューを利用してビジュアル化した情報を比較する予備実験を行った．その結果，今回提案した手法では従来と違う情報を提供できることがわかった．しかし，同時に不必要な単語の表示が確認された．この課題を解決していく必要がある．

今後は，文章中のキーワード抽出精度の向上と，ML-Ask の感情分類の正確性向上を図る．その後，提案手法を組み込んだ実験用のシステムを構築し，実験によって本研究の有用性を検証する．

参考文献

1) 橋本雄太. “近代デジタルライブラリーのためのソーシャルリーディング環境の構築”. 研究報告人文科学とコンピュータ（CH）2014, no. 9 pp 1–5. 2014.
2) Amazon, https://www.amazon.co.jp/, [Accessed: February 15, 2017].
3) 読書メーター, http://bookmeter.com/, [Accessed: February 15, 2017].
4) idpf. “EPUB | International Digital Publishing Forum”, http://idpf.org/epub. [Accessed: February 15, 2017].
5) 片岡えり，天笠俊之, Franck Gass, 北川博之, “EPUB を対象としたソーシャルリーディングシステムにおけるユーザ分析”, DEIM Forum 2015, 2015.
6) ipdf. “EPUB Canonical Fragment Identifiers 1.1”, http://www.idpf.org/epub/linking/cfi/epub-cfi.html, [Accessed: February 15, 2017].
7) 王淼，大塚隆弘，榎原博之. ”アノテーション機能を備えた文献評価システムの構築”, 研究報告電子化知的財産・社会基盤（EIP）2011-EIP-53(19), pp 1-7, 2011.
8) Daniel Steinbock. “TagCrowd,”, http://tagcrowd.com/faq.html#whatis, [Accessed: February 15, 2017].
9) MeCab: Yet Another Part-of-Speech and Morphological Analyzer http://taku910.github.io/mecab/, [Accessed: February 15, 2017].
10) 飯田龍，徳永健伸. “談話の顕現性を考慮した重要語抽出とその応用”. 研究報告自然言語処理（NL）, pp 1-8, 2009.
11) Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, “Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -”, In Proceedings of The Conference of the Pacific Association for Computational Linguistics (PACLING-09), September 1-4, Hokkaido University, Sapporo, Japan, pp. 223-228, 2009.
12) Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, “A System for Affect Analysis of Utterances in Japanese Supported with Web Mining”, Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 21, No. 2 (April), pp. 30-49 (194-213), 2009.