

閲覧中 Web ページのブロック分割に基づく 書籍推薦のためのブラウザ拡張機能

岡井 義宗[†] 大園 忠親^{††} 新谷 虎松^{††}

名古屋工業大学情報工学科[†] 名古屋工業大学大学院情報工学専攻^{††}

1 はじめに

閲覧中の Web ページに関連した書籍を推薦するシステムの実現が望まれている。本研究では、閲覧中の Web ページに関連した書籍の推薦を実現するための新たなシステムを開発した。任意の Web ページ閲覧における、書籍の検索作業の手間を省くことを目的とした。本システムの特長は、1) ブラウジング中の透過的な推薦、2) Web ページ上の複数コンテンツへの考慮、3) クエリフリー（クエリが不要）、の3点である。本稿では書籍推薦のためのブラウザ拡張機能の実現方法について述べる。

2 ブラウザ拡張機能を用いた閲覧中 Web ページにおける書籍推薦

2.1 ブラウザ拡張機能を用いた透過的な推薦

推薦書籍を提示する際、ブラウジング中の Web ページとは別に、推薦書籍を表示するウィンドウやタブを展開しなければならないという問題がある。そこで本研究では、ブラウザ拡張機能を用いた書籍推薦を行う。本システムにおけるブラウザ拡張機能とは、ウェブブラウザ「GoogleChrome」を拡張するための機能である。本機能では、閲覧中の Web ページに直接影響を与えることなくシステムの実行が可能である。そのため本機能を用いることで、ユーザによる検索作業等の閲覧行動を阻害しない、透過的な推薦システムの実現が可能となった。

2.2 ブロック分割による複数コンテンツ対応

本研究では複数コンテンツを含む Web ページにおいても推薦を可能とするため、Web ページをコンテンツ毎にブロック分割する。佐野ら [1] の研究では、Web

A Browser Extension for Block Segmentation-based Book Recommendation in Browsing Web Pages

[†]Yoshimune OKAI, ^{††}Tadachika OZONO and ^{††}Toramatsu SHINTANI

[†]Dept. of Computer Science, Nagoya Institute of Technology

^{††}Dept. of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology.

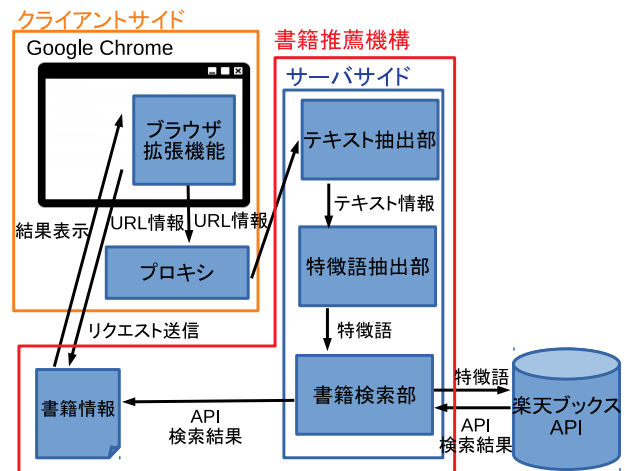


図 1: システム構成図

ページを細分化ブロックという単位まで分割した後に、Web コンテンツの見出しとなるようなブロックに着目して細分化ブロックを結合することにより、Web ページを意味的にまとまりのある単位へと分割する手法を提示している。

本研究ではブロック分割の手法として、閲覧中 Web ページから HTML ソースを取得し、HTML ソースに含まれる DOM 木構造を利用した。Web ページは Document オブジェクトによって表されており、ブラウザが Web ページにアクセスし HTML ソースを解析すると、文書の内容を表すオブジェクトの木構造が構築される。構築された木構造からタグごとにテキストを抽出することで、Web ページのブロック分割を可能とする。

2.3 クエリフリーな書籍推薦機構

本システムでは、検索クエリが不要なシステムを実現するための書籍推薦機構が必要となる。本機構では、分割されたブロック毎に文字列のみを抽出したテキスト情報を利用する。そのため Web ページの特徴抽出には、抽出したテキスト情報を単語毎に分割し、ベクトル化する必要がある。

テキスト解析の流れとして、まずテキスト情報に対

して形態素解析を行う。その後、解析した各形態素に対して TF-IDF 特徴量を計算する。

まずテキスト情報を単語毎に分割し、品詞が名詞である単語のみを抽出する。その後、各名詞に対する TF-IDF 値を求める。今回は各ブロックにおける、TF-IDF 値の大きい上位 5 単語をそのブロックでの特徴語とし、書籍検索のための検索クエリとする。

本システムでは、書籍検索のために外部公開されている「楽天ブックス書籍検索 API」を利用する。この API は、インターネットサービスを提供する楽天株式会社が発行しているサイト「Rakuten Developers」にて提供されている API のひとつであり、書籍のタイトルや著者名、出版社名などに対してクエリ検索が可能である。検索結果として、書籍に関する情報が JSON 形式で返ってくるため、その中から必要な情報を抽出する。今回抽出するのは、「書籍のタイトル」、「著者名」、「出版社名」、「商品内容」、「商品画像」、「商品レビュー平均」、「楽天ブックスジャンル ID」である。これらの要素を用いて、本システムにおける関連書籍の情報提示を行う。

また、各書籍のジャンル名を獲得するために、同サイトにて提供されている「楽天ブックスジャンル検索 API」を利用する。この API では、「楽天ブックスジャンル ID」を検索クエリとして渡すことで、楽天ブックスにて分類されている書籍ジャンル名を獲得することができる。

取得した書籍に関する要素を、本システムにおける書籍情報とし、ユーザに推薦書籍として提示する。本システムの特筆すべき点として、Web ブラウザに表示されている内容を含むブロックを用いた推薦を行うことで、ユーザが閲覧中のコンテンツ領域に特化した推薦が可能になる点が挙げられる。

2.4 システム構成図

本システムにおける構成図を図 1 に示す。本システムでは、ユーザのリクエスト処理などを行うクライアントサイドと、Web ページを解析し、関連した書籍情報を取得するサーバサイドに分けられる。ユーザが Web ページに遷移した際に、閲覧中 Web ページの URL 情報がプロキシへ送信される。これは、Google Chrome における拡張機能の開発において、クライアントサイドから外部に直接アクセスすることが制限されている。そのため、本拡張機能が、直接サーバサイドへ通信するためのプロキシを導入した。本プロキシを経由することで外部と通信可能になる。その後、プロキシで受



図 2: 実行画面¹

信した URL 情報はサーバサイドであるブロック分割部に送信され、URL 情報から Web ページの HTML ソースを取得し、Web ページのブロック分割処理を行う。分割された各ブロックは、ブロック情報としてテキスト抽出部に送信され、ブロック毎にテキスト情報を取得する。また、テキスト情報から抽出した特徴語は書籍検索部へと送信され、特徴語を楽天ブックス API に対する検索クエリとする。得られたクエリにより楽天ブックス API を用いて書籍情報を検索する。

3 システム実行例

本システムの実行画面を図 2 に示す。本システムでは閲覧中の Web ページ上に表示させるため、ポップアップ機能を利用して関連書籍の提示を行う。ページ遷移後、本ブラウザ拡張機能は透過的な書籍推薦のために、自動的に推薦候補を検索する。ユーザが推薦結果を表示するためのボタンをクリックすると関連書籍が表示され、ユーザは検索質問を入力する必要がない。また、ブロック毎に推薦書籍を提示するための本棚が自動的に生成され、表示される。

4 おわりに

本稿では、閲覧中 Web ページにおいて、コンテンツ別に関連した書籍を推薦し、ブラウザ拡張機能を用いて提示するシステムの開発について述べた。ブラウザ拡張機能を利用することにより、検索作業等の Web ページ閲覧行動を阻害することなく関連書籍の提示が可能となった。

謝辞 本研究の一部は JSPS 科研費 JP15K00422, JP16K00420 の助成を受けたものです。

参考文献

- [1] 佐野博之, 白松俊, 大園忠親, 新谷虎松: Web ページ分割のための決定木学習を用いたタイトルブロック抽出, 電子情報通信学会論文誌, vol.J95-D, no.4, pp.909-918 (2012).

¹<https://qiita.com/satoken0417/items/479bcd91cff2634ffb1>