

大学図書館における電子化を想定し可視化を用いた 検索システムの研究

小川 友暉[†] 相場 亮[‡]

芝浦工業大学大学院電気電子情報工学専攻[†] 芝浦工業大学システム理工学部電子情報システム学科[‡]

1. はじめに

近年、電子図書館が注目を集めているが、キーワードを主体として本を探しているため、実際の本が並ぶ図書館のように本棚を歩きまわって本を探し出す、いわゆるブラウジングを行うことが出来ない。ブラウジングは、関心のある内容が曖昧でも本を探せるという利点がある。これを電子図書館上で実現することによって、本が探しやすくなる。本研究は、本棚の整理に用いられる図書館分類で関心のある内容に近い本を探し、トピックマップを用いて可視化することで電子図書館上に、ブラウジングの利点を実現することを目指す。

2. 図書館分類法の特徴と利点

図書館分類の特徴の一つは、用途ごとに分けられていることである。例えば、キーワードを Javascript で検索した場合の分類の例を図 1 に示す。図書館分類の一つである日本十進分類法 [1] (以下 NDC とする) の分類番号を左側、検索された本のタイトルを右側に表す。

007.6	JavaScript プログラミング&サンプル
547.48	JavaScript 辞典: HTML5 対応
694.6	Web 技術者のための HTML5+JavaScript で作る Android アプリ
798.5	JavaScript ゲーム制作教科書

図 1. NDC の例[2]

図 1 の例では、007.6 はプログラミング、547.48 はクライアントサーバーシステム、694.6 はスマートフォン、798.5 はコンピューターゲームを表す分類番号である。このことから、同じ Javascript の本でも用途が違ってくる。この分類番号があるため、例えば、利用者がゲームを作る目的で、Javascript の本を探す場合に、同じ 798.5 の

分類番号を持った異なる本が見つかるといった利点がある(図 2)。

798.5	Python ゲームプログラミング入門
798.5	Windows ゲームプログラミング
798.5	Unity ではじめるゲームづくり

図 2. 798.5 の例[2]

したがって、図 2 のように、直接キーワードで調べなくても、図書館分類に沿って本棚に本が並べられているおかげで、利用者の関心のある内容の本を見つけることが出来る。

しかし、逆に Web 上で図書館分類を用いて本を探す場合には、分類番号の入力が必要となるので、図書館に詳しくない利用者が使用するのは困難である。

3. トピックマップ

トピックマップ[3]は、利用者のもつ概念体系に合わせて情報を分類、整理するための国際規格 (ISO/IEC 13250) であり、情報の見つけやすさに重点を置いている。

トピックマップの構成要素は、主題を表現する「トピック」、主題間の関係を表す「関連」、主題と主題に関連する情報リソースを結びつける「出現」の 3 つからなる。

4. 関連研究

本研究に類似した研究として 2 つ挙げる。

1 つ目は、國岡ら[4]が、論文検索において、調査対象に関する知識を十分に持たない利用者を支援するシステムを提案している。國岡らのシステムでは、調査対象に関係するシソーラスなどの知識を図として可視化することで、論文検索の支援を行っている。

2 つ目は、常川ら[5]が、NDC を用いた、読書傾向が似ているユーザー発見手法を提案している。類似の読書傾向を持つ読者を発見する手法として、“NDC ツリープロファイリング”を提案している。この手法で生成されるユーザー同士のプロファイルを比較することで、読書傾向の類似したユーザーを発見する。

A search system using visualization for digital libraries in universities

Ogawa Yuki[†], Aiba Akira[‡]

[†]Electrical Engineering and Computer Science, Graduate School of Engineering and Science, Shibaura Institute of Technology

[‡]Department of Electronic Information Systems, College of Systems Engineering and Science, Shibaura Institute of Technology

5. 提案手法

本研究は、利用者が関心のある本を探す支援手法を提案する。図2の例から、図書館では図書館分類によりブラウジングが実現できている。そこで、本手法では図書館分類を用いて情報を集め、利用者に提示することでブラウジングの利点の実現を目指す。本に関する情報は書誌情報であり、書名、ISBN、NDCの分類番号である。

提案手法の流れは以下の通りである。

- ① 利用者がキーワードを入力する。
- ② 本システムがキーワードに該当する本を収集する(タイトルの部分一致)
- ③ キーワードを用いて収集した本と同じ分類番号を持つ本も収集する
- ④ コサイン類似度を用いてフィルタリングを行う(フィルターの閾値は可変)
- ⑤ フィルタリングされた情報からトピックマップを生成し、表示する。

5.1. トピックマップの使用方法

トピックマップの要素を以下のように定義する。

- 中心ノード : 利用者が入力したキーワード
(トピック)
- 分類ノード : NDC (トピック)
- 本ノード : 収集された本 (出現)

5.2. コサイン類似度

コサイン類似度とは、ベクトル空間モデルにおいて、文書同士を比較する際に用いられる類似度計算手法である。文章ベクトル x , y があるとき、次で求められる。

$$\cos(x,y) = \frac{x \cdot y}{|x| \times |y|}$$

本手法では、Mecab[6]で形態素解析を行い、名詞だけを抽出したものを文章ベクトルとする。

6. 実験

提案手法を基に実験システムを実装し、被験者に利用してもらい、インタビューすることで評価する。開発環境等は以下の通りである。

- Java (システム開発言語)
- TouchGraph[7] (トピックマップの描画処理)
- Selenium[8] (検索後に OPAC への誘導処理)
- 日本十進分類表[9]
- 芝浦工業大学の OPAC データの一部[2] (書誌情報)

システム構成を、図3に示す。利用者は、キーワード入力や分類ノード・本ノードをクリックすることで、OPACの詳細情報を見ることが出来る。

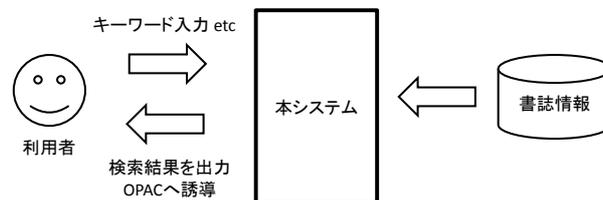


図3. 実験システムの構成図

実験方法としては、利用者に予め、インタビューを行い、図書館の利用経験、専門分野、興味がある分野を聞き、その後、実験システムを利用して、被験者の専門分野ではない興味がある本を探してもらう。本研究は、ユーザービリティの影響を受けなくするために、システムの操作等は筆者が行い、被験者には出力結果のみを提示する。

7. 現状

システムの実装は終了し、予備実験として利用してみたが、以下の問題がある。

- ① 出力されるノード数が多すぎる。
- ② ノードの位置が密集してしまっている。

①の原因は、コサイン類似度で算出した値の同じ本が膨大な数得られることにある。②の原因は、TouchGraphのノードの座標算出アルゴリズムによるもの影響もあると考えられる。

参考文献

- [1] 緑川 信之, 本を分類する, 勁草書房, 1996
- [2] 芝浦工業大学 図書館 WebOPAC. <https://www.lib.shibaura-it.ac.jp/webopac/topmnu.do>
- [3] 内藤 求, トピックマップ入門, 東京電機大学出版局, 2006
- [4] 國岡 崇生, 田村 友紀, 山崎 文枝, 堀内 美穂, 坂内 悟. JST シソーラス map JST 辞書の可視化による効果的な検索語の発見. 情報管理. 2012, vol. 55, no. 9, p. 662-669.
- [5] 常川 真央, 松村 敦, 宇陀 則彦. 日本十進分類法を用いた類似読者発見手法. 情報メディア研究 12(1), 42-51, 2013
- [6] MeCab. <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- [7] TouchGraph. <http://sourceforge.net/projects/touchgraph/>
- [8] Selenium. <http://www.seleniumhq.org/>
- [9] 日本十進分類表. <http://www.zuki-ebetsu.jp/ndc/>