

電子教科書の学習分野を考慮したインターネット検索の提案

Study Field Judgement Function in Internet Searching for Digital Textbooks

大国 航^{*1}, 下倉 雅行^{*1*2}, 島袋 舞子^{*1}, 兼宗 進^{*1}, 村上 晴美^{*2}
 Wataru OGUNI^{*1}, Masayuki SHIMOKURA^{*1*2}, Maiko SHIMABUKU^{*1}, Susumu KANEMUNE^{*1},
 Harumi MURAKAMI^{*2}

*1 大阪電気通信大学

*1 Osaka Electro-Communication University

*2 大阪市立大学

*2 Osaka City University

Email: ht12a009@oecu.jp

あらまし：電子教科書を用いて学習する際、学習者が内容をより深く学習するためにはインターネット検索が有用であると考えられる。しかし、インターネット検索の結果をそのまま利用しようとすると、どの結果が正しいかを判定することが初学者にとっては困難であると推測される。そこで、検索語が含まれる電子教科書の段落と検索結果の一部を分野判定し合致したものを提供することで、学習者の検索に対する負担を軽減することを考えた。ここでは、この検索機能について説明する。

キーワード：電子教科書、電子書籍、情報検索、分野判定

1. はじめに

近年、小学校から高等学校まで電子教科書の利用が始まっている。タブレット端末などで容易に閲覧することが可能で、用語の解説機能が搭載されているものもある。しかし解説機能で提供されること以上のことを学習しようとする場合や、解説機能に搭載されていない用語を調べたい場合にはインターネット検索を利用することになる。インターネット検索を利用する場合、ブラウザを別途立ち上げ、検索サイトで調べたい単語をインターネット検索する。しかし、学習途上の学習者にとって、単語のみで検索した結果からどれが正しいのか判断できず、全く違う分野の言葉を見てしまうこともありうる。

そこで我々は、電子教科書から検索した際、検索結果と電子教科書の分野が同じであるものを選び、提供する機能(分野判定検索機能)を考案した。

2. 分野判定検索機能の設計

一般的にインターネットの検索結果としてどのような結果が得られるかは検索エンジンが持つランキングシステムに依存する。そのため、教科書に載っている用語であっても、期待したものと大幅に異なる結果が得られることもある。そこで、検索結果の分野を判定し、検索語が含まれる電子教科書の段落の分野と合致したもののみを提供することにより、学習分野に適した検索結果を提供することが可能になると考えた。流れとしては次のようになる。

- (1) 分野判定のための辞書(分野判定辞書)を作成
- (2) 電子教科書の本文を、分野判定辞書を用いて段落ごとに分野判定
- (3) 電子教科書の調べたい語を検索
- (4) 検索結果のサイトから抜粋された文章(スニペット)を、分野判定辞書を用いて分野判定

- (5) 検索語が含まれる段落の分野と、スニペットの分野を比較
 - (6) 分野が合致した Web サイトを検索結果として学習者に提示
- この流れを図にしたものが図1である。

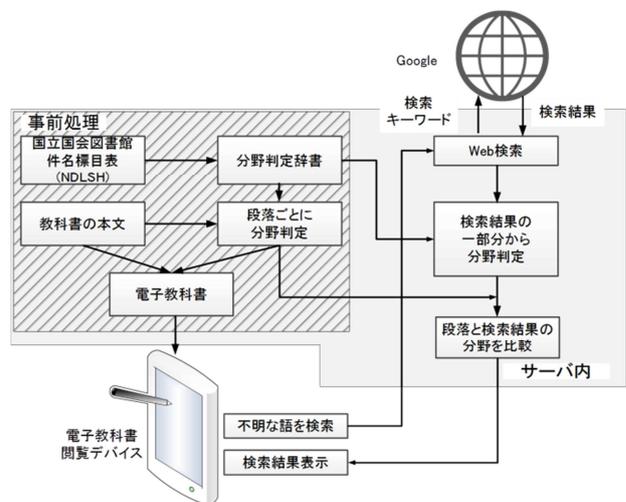


図1 分野判定検索機能の流れ

分野判定辞書を作成するため、単語と分野の対応を示すデータとして国立国会図書館件名標目表(NDLSH)⁽¹⁾を利用した。NDLSHには、件名標目、標目よみ、ID、同義語、上位語、下位語、関連語、注記、分類記号(NDLC)、分類記号(NDC9)、参照(LCSH)、参照(BSH4)、出典(BSH4)、出典、編集履歴、作成日、最終更新日が含まれる。このうち、目録を検索する手がかりとして用意されている件名標目と分類記号(NDC9)を利用することとした。分野判定を行う時には、判定対象の文章を文節によって

区切ることで単語に分割する。ここから、件名標目も同様の処理で単語に分割し、件名標目が持つ分類記号 (NDC9) を取り出す。次に単語の出現頻度 Term Frequency⁽²⁾ をカウントし、この数によって重みを決める。ここで得られた単語、分類記号 (NDC9)、重みをまとめたものを辞書とした。

電子教科書の検索語が含まれる段落を分野判定した結果と、スニペットを分野判定した結果を比較し、合致するものを提示することで、学習内容と関連する Web サイトのみを閲覧することが可能になる。

3. 分野判定検索機能の評価実験

3.1 評価方法

分野判定検索機能を用いて、どのような結果が得られるのかを調べた。テストデータとして、小学6年生社会科の『小学社会6年上⁽³⁾』、小学6年生理科の『わくわく理科6⁽⁴⁾』、高等学校情報科の『社会と情報⁽⁵⁾』の教科書を利用した。

教科書中の用語(10語)をインターネット検索する。人間とシステムが検索した用語が含まれる文章の内容と、検索結果(10件)のスニペットの内容が同じ分野であるかを正誤判定する。判定結果から、情報検索の評価によく用いられる⁽²⁾、適合率(式1)、再現率(式2)、F値(式3)を計算して、各教科の平均値を算出する。なお、検索する用語は理科と情報科は教科書の索引ページから選択し、社会科は教科書に索引ページが存在しなかったため、学習指導要領⁽⁶⁾に記載されている用語から選択した。人間の判定は大学生5名が判定をし、多数決をとり1つの判定とした。

$$P = \frac{\text{取得された正解要素数}}{\text{取得された要素数}} \quad (\text{式1})$$

$$R = \frac{\text{取得された正解要素数}}{\text{正解要素数}} \quad (\text{式2})$$

$$F = \frac{2PR}{P+R} \quad (\text{式3})$$

3.2 結果

各教科から10個ずつ、合計30個の検索語について適合率、再現率、F値を調べた。その結果、社会科から「廃藩置県」「武家諸法度」「日清戦争」、理科から「化石」「断層」は適合率、再現率、F値すべてが0.7以上を示し、よい結果が得られた。

全体をみるために各教科の平均値をまとめたものが表1である。検証した3つの科目のうち、社会科と情報科の適合率、再現率が低くなってしまっている。適合率が低い原因としては、分野判定を行う際に幅広く分野を認めていたため、本来であれば分野が異なるはずのサイトであっても同じ分野と判定してしまったことである。このため、表示すべきでない検索結果も表示してしまい、適合率が低下してしまった。

次に、再現率が低い原因としては、検索結果の分

野判定時に、スニペットが短すぎるとどの分野にも分類されないことで検索結果として提供されないことである。これにより、表示すべき検索結果が表示されないことで、再現率が低下した。特に辞書サイトや用語解説サイトなどで発生している現象である。

表1 分野判定検索機能の適合率、再現率、F値

教科	適合率	再現率	F値
社会科	0.36	0.57	0.43
理科	0.61	0.80	0.62
情報科	0.28	0.43	0.32

4. おわりに

本研究では、電子教科書で利用することが可能な学習分野を考慮したインターネット検索機能を提案した。学習分野を考慮したインターネット検索機能は、検索したい語が含まれる電子教科書の段落と、検索結果のスニペットを分野判定して、それぞれを比較し合致する Web サイトのみを検索結果として提供することで、学習を支援することを目的とした。

性能評価を行い、用語によっては正しく分類されて有効であることがわかった。しかし、今回提案した分野判定検索機能には、判定する文章によっては多くの分野が該当してしまう、判定する文章が短すぎると分野判定ができない、という問題点が見つかっている。このため、まだ学習の支援としては不十分である。今後、分野を判定するための辞書の作成方法の見直しや分類方法の見直しが課題として挙げられる。

参考文献

- (1) 国立国会図書館 - National Diet Library .
<http://www.ndl.go.jp/jp/library/data/bunruikenmei.html>
- (2) Christopher D.Manning, Prabhakar Raghavan, Hinrich Schutze 著, 岩野和生, 黒川利明, 濱田誠司, 村上明子 訳: 情報検索の基礎, 共立出版(2012)
- (3) 清水毅四郎ほか: 小学社会6年上, 日本文教出版株式会社(2014)
- (4) 大隅良典ほか: わくわく理科6, 新興出版社啓林館(2014)
- (5) 赤堀侃司ほか: 社会と情報, 東京書籍株式会社(2012)
- (6) 文部科学省: 現行学習指導要領・生きる力, http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/sy0/