

# 電子教科書で利用可能な分野判定検索の提案

下倉 雅行<sup>1,2,a)</sup> 大国 航<sup>1,b)</sup> 島袋 舞子<sup>1,c)</sup> 兼宗 進<sup>1,d)</sup> 村上 晴美<sup>2,e)</sup>

**概要：**電子書籍の普及が進んでおり、端末では国語辞典などで語句の意味を調べながら読書を行う機能が提供されている。今後普及が予想される電子教科書においては、一般的な語句の意味にとどまらず、学習内容に関連する Web 上の情報源にアクセスしてより深い学習を行うことが有効と考えられる。しかし、単純に Web 検索の結果をすべて表示すると、どの検索結果が正しいのかを判定することが学習を始めようとする学習者には困難であることが予想される。さらに、学習内容と異なる内容の Web ページを閲覧してしまう可能性も存在する。そこで我々は電子教科書に適した Web 検索手法を考案し試作を行った。検索したい語が含まれる電子教科書の段落を分野判定したものと、検索結果のスニペット部分を分野判定したものが合致するページのみを提供することで、学習中の文脈に対応した Web 検索が可能になる。

**キーワード：**電子教科書, 電子書籍, 情報検索, 分野判定

## Field Judgment Function in Web-Searching for Digital Textbooks

SHIMOKURA MASAYUKI<sup>1,2,a)</sup> OGUNI WATARU<sup>1,b)</sup> SHIMABUKU MAIKO<sup>1,c)</sup> KANEMUNE SUSUMU<sup>1,d)</sup>  
MURAKAMI HARUMI<sup>2,e)</sup>

**Abstract:** The worldwide popularity of digital books continues to increase. One useful function is that they display word meanings with an embedded dictionary function. Such web-searching functions are becoming very effective and more popular for learners. However, students often have difficulty selecting correct meaning from multiple meanings on the screen. This may result in misunderstandings and confusions. Therefore, our presentation enables context-based web-searching in digital textbooks by proposing a new classification function that determines a word's fields from the textbook in which it appears and displays search results limited to specific fields.

**Keywords:** digital textbook, digital book, information retrieval, field judgment

### 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でき、さら

に動画などのコンテンツを含むことが可能である。また、電子書籍には国語辞書や英和辞書などの辞書により、言葉を調べる機能(図1)が付属している。これは言葉の意味が出てくるものであり、専門的な用語に関する答えを得られるものではない。電子教科書では、通常の国語辞書や英和辞書などから得られない用語も出てくることもあり、それらの用語についてはインターネット検索を利用することとなる。インターネット検索を利用するとすれば、ブラウザを別途立ち上げ、検索サイトで調べたい単語を検索することとなる。学習途上の学習者にとっては、この未知の用語を検索する、という行為はかなり敷居が高く、検索結果が得られたとしても、どれが正しいのか理解できず、全く

<sup>1</sup> 大阪電気通信大学  
Osaka Electro-Communication University, 1130-70, Kiyotaki, Shijonawate, Osaka 575-0063, Japan

<sup>2</sup> 大阪市立大学  
Osaka City University, 3-3-138, Sugimoto, Sumiyoshi, Osaka 558-8585, Japan

a) masayuki@karekinada.com

b) ht12a009@oecu.jp

c) shimabuku.m@gmail.com

d) kanemune@acm.org

e) harumi@media.osaka-cu.ac.jp

違う分野の言葉を見てしまうこともありうる。学習するための情報の難易度となじみのない言葉の意味の理解を支援するシステム [1] もあるが、wikipedia に限定しており、一般的な Web ページには対応していない。

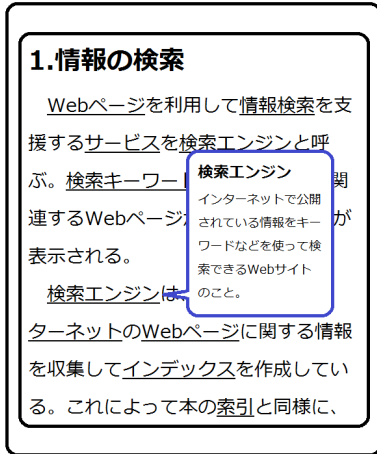


図 1 辞典機能を備えた電子書籍の画面例

検索支援として、(1) ブックマーク共有により検索支援するもの [2], (2) 検索ログを利用したニーズを抽出 [3], (3)HTML タグによるクラスタリング手法 [4], (4) 小規模カテゴリ構造を構築する検索支援 [5] などが存在している。しかし、学校教育の場では、(1)(2) はブックマークがない状態や検索ログがない状態も考慮しなければならないため、利用が困難となる。また、(3)(4) は Web ページ自体の分類となり、教科書の分類と合致するとは限らない。

そこで、我々は、電子教科書から直接検索でき、図 2 のように検索結果と電子教科書の分野が同じであるものを選別できる機能を考案した。ここでは、この検索機能の実装について説明する。

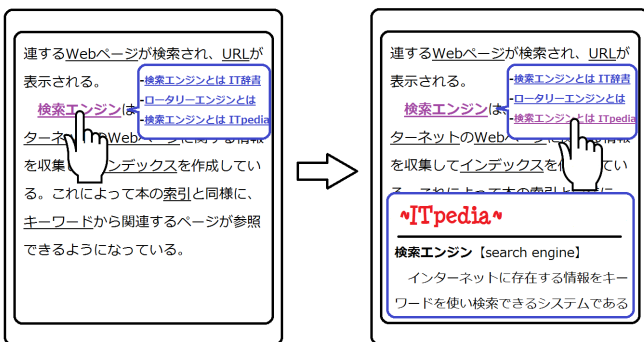


図 2 分野判定検索機能付電子教科書

## 2. 電子教科書に求められる機能

電子教科書に求められる機能は、日本教育情報化振興会 (JAPET) が 2010 年 11 月に「デジタル教科書の位置づけと機能」[6] という資料の中でデジタル教科書に必要な機能としてまとめている。これを表にしたものが表 1 である。

表 1 デジタル教科書に必要な機能

No.	機能	概要
1	マルチメディア提示機能	教科書に即した内容を、音声、写真、映像、アニメーションなどを利用してわかりやすく提示する
2	教材カスタマイズ機能	教師の指導方法や学習者の特性に応じて、教材を選択または編集できる
3	関連教材・資料へのリンク機能 (内部・外部)	学習内容に関連教材・資料等にアクセスできる。
4	内容の自動更新機能	統計値、地図、地名、制度などを常に最新の状態に保つ。
5	学習習熟度に応じた演習問題提供機能	教師の指示で個々の学習者に適した演習問題を設定できる。
6	学習者データの収集・分析機能	演習問題等の実施結果を収集し、教師が分析しやすいかたちで提示する。
7	学習者の入力インタフェース機能 (認識機能)	手書きによる筆算や計算の途中過程も認識できる。
8	コミュニケーション機能	学校内の児童生徒間コミュニケーション、学外とのコミュニケーション、宿題や課題の提出、家庭とのコミュニケーション
9	各種ツール機能	資料まとめ・発表、資料の協働製作、音楽、美術等の創作、辞書、事典

電子教科書が活用される場面として、同じ JAPET の資料の学習者用のデジタル教科書活用場面例の中の 1 つに「(3) 個々にあるいはグループで主体的な、探求的な学習」とある。また、この具体的な例として「インターネットなどによる調べ学習」が挙げられている。しかしながら、求められる機能の中には、この調べ学習に利用できる機能が挙げられていない。

学習者が電子教科書を利用して予習や復習をする際にわからない単語があると、そこで辞書やインターネットを利用して調べることとなる。辞書に載っている単語のみであ

れば問題はないが、辞書にない単語をインターネットの検索エンジンを利用して調べるとなると、検索結果が正しいかどうかの判別もできない。そこで、現在挙げられている電子教科書に求められる機能に加えて、検索をサポートする機能が必要であると考え。

ここでいう検索をサポートする機能は、教科書の各節や項の関連する分野に検索結果を限定し、大幅に異なった分野を出さないようにする機能(ここでは分野判定検索機能と呼ぶ)である、と考えている。

### 3. 分野判定検索機能の設計と実装

#### 3.1 分野判定検索機能の概要

Web 検索結果は、どのような結果が得られるかはわからない。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がよくある。そこで、検索結果の分野を判定し、検索語の分野が合致したのみを提供することで、この大幅に異なる結果のために起こりうる、間違えた学習を減らすことが可能になると考えた。流れとしては次のような形になる。

- (1) 電子教科書上で意味のわからない語句をクリック(またはタッチ)
- (2) 語句が含まれる段落(または節や項)から分野を判定
- (3) 語句を Web 検索にかけて得られた結果の一部(スニペット部分)を分野判定
- (4) 元々の語句の分野と同じ分野の検索結果のみ提供

#### 3.2 分野判定検索機能の設計

分野判定検索機能を実現するための流れを整理する(図3)。

- (1) 分野判定のための辞書(分野判定辞書)の作成
- (2) 電子教科書の本文を段落ごとに分野判定
- (3) 電子教科書の本文を HTML に変換(語句ごとの検索機能を利用するためのリンク作成)
- (4) 検索エンジンを利用した検索
- (5) 検索結果を分野判定
- (6) 分野が合致する結果のみ表示

分野判定辞書を作成するためには、単語と分野の対応を示すデータが必要である。今回の試作では、国立国会図書館件名標目表[7](NDLSH)を利用した。件名から得られたキーワードと NDC 分類記号を抜き出す。さらに、キーワードの出現頻度 TF(Term Frequency)[8]をカウントし、重み付けを行えるようにする。これにより得られた辞書を用いて、分野を判定する。

検索エンジンへのアクセスには、検索エンジンが提供する API を利用した。今回利用したのは、Google Custom Search API である。検索結果の分野判定には、検索結果から得られるスニペットを利用する。

電子教科書の本文の、検索語の含まれる段落を分野判定

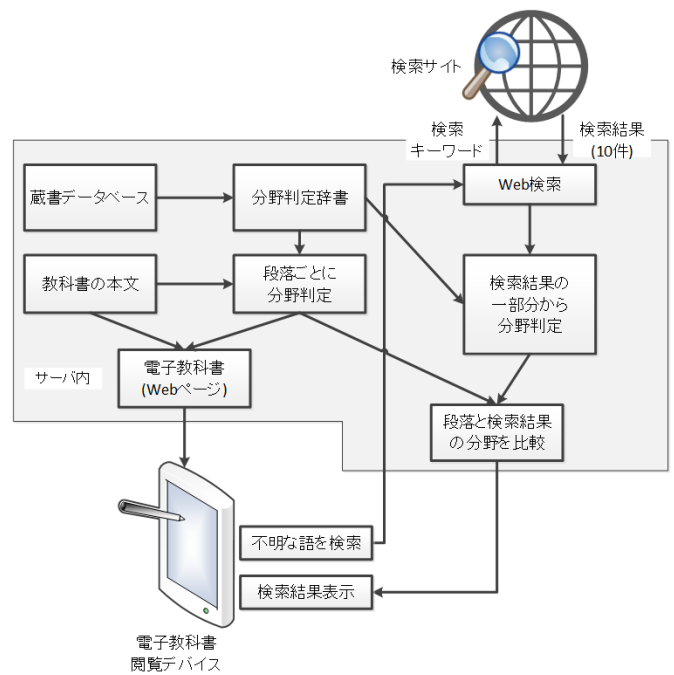


図3 分野判定検索機能の流れ

した結果と、スニペット分野判定した結果が合致するもののみを提示することで、学習内容に近い検索結果だけを提示することが可能になる。

#### 3.3 分野判定検索機能の実装

分野判定辞書は、国立国会図書館件名標目表を元に作成した。国立国会図書館件名標目表の件名の語を KAKASI で分割し、漢字、カタカナ、英語を辞書のキーワードとし、そこに含まれる NDC 分類記号を抽出する。このとき、同じキーワードが複数の NDC 分類記号を持つ場合は、キーワードごとに複数の NDC 分類記号と出現回数を持たせている。

分野判定の方法として、文章を KAKASI(kanji kana simple inverter)[9]を利用して分割し、各語ごとに分野判定辞書のキーワードに合致するかを調べ、合致した場合に、NDC 分類記号ごとに重みに応じた値をカウントする。この結果、最も多かった NDC 分類記号が、その文章の分野と判定している。

分野判定検索機能を利用する電子教科書は、wiki をベースとした QHM(Quick Homepage Maker)[10]上に作成されたものを想定している。この電子教科書の本文に対して、事前に段落ごとの分野判定を行い、段落と分野の対をデータとして保持している。そして、電子教科書の本文を、wiki の形から HTML に変換し、この HTML に各語ごとのリンクを作成する。このリンクをクリックすることで、分野判定検索機能呼び出す。このリンクには、検索したいキーワードと段落番号との対を含むようにしている。

分野判定検索機能は、電子教科書の本文からのリンクで

渡されたキーワードと段落番号から、キーワードを検索エンジンに渡す。検索エンジンから得られた結果のうち、最初の10件のみを利用する。検索結果には、Webページのタイトル、URL、スニペットなどがあるが、このうち、スニペットに含まれる文章を分野判定に利用する。スニペットから判定された分野と、段落番号から得られる分野とを比較し、合致するもののみ検索結果として表示する。

#### 4. 分野判定検索機能の評価実験

実際に分野判定検索機能を用いて、どのような結果が得られるかを調べてみた。テストデータとして、高等学校情報の「社会と情報」の教科書を利用し、理論に関するある節にある語について、各10個ずつ検索した結果を調べた。検索結果に対して、その語が含まれる項の分野と合致しているかどうかを、人とソフトともに正しい(人判定○機械判定○)、人とソフトともに間違い(人判定×機械判定×)、人は正しいがソフトは間違い(人判定○機械判定×)、人は間違いソフトは正しい(人判定×機械判定○)、と4つに分けてまとめたものが表2である。

表2 分野判定検索結果判定

検索語	人判定		機械判定	
	○	×	○	×
マスメディア	5	0	1	4
コンピュータ	1	2	3	4
データ	3	2	1	4
メディア	3	2	0	5
ハードディスク	3	1	1	5
インターネット	3	0	1	6
信頼性	4	0	0	6
ビット	1	8	1	0
情報操作	4	4	1	1
二次情報	5	4	0	1
信憑性	4	1	5	0
著作権	7	0	3	0
暗号化	4	3	3	0
信用毀損	2	0	8	0
被告訴人	3	2	5	0
書類送検	2	4	4	0
データ圧縮	1	2	7	0
著作権侵害	6	1	3	0
メディア変換	5	2	0	3

検索結果から、一般的な言葉(マルチメディアやハードディスク、インターネットなど)は少し関係の無いサイトを排除できることもあるが、ほとんど通常の検索結果と同じである。しかし、一般用語から少し離れた書類送検や二次情報、暗号化などは、不要なものが排除され、必要なものが得られるようになっている。ここから、ここで作成した分野判定検索機能は、一般用語ではあまり効果が得られ

ないが、専門用語では余計なものを排除する効果が得られることがわかった。

#### 5. おわりに

今回作成した分野判定検索機能は、専門用語に対しては、余計なサイトを除外することができた。今回検証したものは高等学校「社会と情報」の教科書における言葉のみだが、学習者がインターネット検索を利用して、より深く学習する際には有用であると考えられる。しかしながら、一般的な語に対しては、通常の検索結果とほぼ同様になり、効果が得られなかった。例えば、「信頼性」のような複数の分野で利用されるような言葉では、うまく検索結果の分野判定に失敗してしまった。この場合はさらに深く検索を行うか、検索語を追加する必要がある。

今後、さらに分野判定辞書の精度を上げるための手法の考案と、正しい検索を得られるような改善を施すことを検討する。また、別の教科や校種の教科書での検証もまだ行っていないため、さらなる有用性の検証をおこなっていく予定である。

#### 参考文献

- [1] 西原陽子, 砂山渡, 谷内田正彦: Webページの難易度と学習順序に基づく情報理解支援システム, 電子情報通信学会論文誌, Vol.J89-D, No.9, pp.1963-1975, 2006.
- [2] 森幹彦, 山田誠二: ブックマークエージェント: ブックマークの共有による情報検索の支援, 電子情報通信学会論文誌, Vol.J83-D1, No.5, pp.487-494, 2000.
- [3] 大久保雅且, 杉崎正之, 井上孝史, 田中一男: WWW検索ログに基づく情報ニーズの抽出, 情報処理学会論文誌, Vol.39, No.7, pp.2250-2258, 1998.
- [4] 折原大, 内海彰: HTMLタグを用いたWebページのクラスタリング手法, 情報処理学会論文誌, No.49, Vol.8, pp.2910-2921, 2000.
- [5] 仲川こころ, 高田喜朗, 関浩之, 可変なカテゴリ構造を用いた文書検索支援手法, 情報処理学会論文誌 Vol.42, No.10, pp.2441-2453, 2001.
- [6] 一般社団法人日本教育情報化振興会: デジタル教科書の位置づけと機能, <http://www.japet.or.jp/jog7cjygu-403/>, 2010.
- [7] 国立国会図書館: 分類・件名 (NDLC, NDLSh など) | 国立国会図書館 - National Diet Library. <http://www.ndl.go.jp/jp/library/data/bunruikenmei.html>
- [8] 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.
- [9] KAKASI project: KAKASI - 漢字→かな (ローマ字) 変換プログラム, <http://kakasi.namazu.org/index.html.ja>, 2014.
- [10] 株式会社 北摂情報学研究所: Quick Homepage Maker 簡単ホームページ作成ソフト, <http://www.open-qhm.net/>