

# シラバスシステムからのキーワード抽出による 図書館蔵書検索システムとの連携

田中 要江<sup>†1</sup> 井上 仁<sup>†2</sup>

現在、大学では様々な学習支援が行われており、学生が情報を入手しやすい環境を整えることは重要となっている。本論文では、シラバスに記載されている教科書や参考書の情報を容易に見つけることができるように、シラバスシステムと図書館蔵書検索システムを連携させたのでその手法を報告する。シラバスシステムへの入力とは従来と同じ方法で、仕様は全く変更されておらず、既存のシラバスシステムのプログラムはほんの数行変更するだけであり、図書館蔵書検索システムには一切手を加えていない。また、対象としているシラバスシステムにおける教科書や参考書の指定は自由記述のため、シラバスからキーワードを抽出し簡単なテキスト処理を行い、図書館蔵書検索システムに渡す方式を採用した。

## A method for connecting between a syllabus system and a library catalog system by extracting keywords

TOSHIE TANAKA<sup>†1</sup> HITOSHI INOUE<sup>†2</sup>

A variety of learning support has been performed in university, and it is getting to be more important to provide environments for students to get information easily. In this paper, we present a method for connecting between a syllabus system and a library catalog system to find information about textbooks and reference books written in syllabus. We have changed only a few lines in the existing syllabus system and have not changed the library catalog system at all. Since our syllabus system allows free descriptions for textbooks and reference books fields, we have adopted a method for extracting keywords from the syllabus system by textual processing.

### 1. はじめに

現在、大学ではさまざまな学習支援が行われており、学習に必要な情報を入手しやすい環境を構築していくことも重要な課題となっている[1]。その機能を果たす役割を担っている部署の一つとして図書館があり学習支援の内容も社会の流れに合わせて充実させていく必要がある。また、学習支援を行うにはシステム環境を整えることも欠かせないものとなっており、今までのさまざまな取組みがされてきた結果、充実したデータ量が蓄積されたと言える。これからは、その情報を利用者が活用しやすい環境にしていくことも重要視されており、その方法としてシステム連携も一つの方法であると考えている。現在は、大学内にさまざまな情報（データベース）が散在しており、学生に有用な情報があるにもかかわらず、その情報を見つけれない利用者が多数存在しているという現状がある。そこで、これらの問題をシステム連携により解決したいと考えている。本稿では、システム連携の一つとしてシラバスシステムと図書館蔵書検索システムの連携を行ったので、その内容を報告する。

シラバスには、授業概要や学習目標、教科書、参考書等

が記されており、履修登録時や授業期間中に参照される。図書館では学習支援のひとつとしてシラバスに指定しているテキスト・参考書を揃えている。しかしながら、図書館が図書館に配備されているにも関わらず学生がその事実を知らないこともあると思われる。そこで、その事実を学生に周知するため、学生の情報探索行動に合わせて簡単に検索できるようにする必要がある。

シラバスに指定された教科書や参考書の所在をシラバスから探索できるようにしている先行研究としては、岩手大学[2]、佐賀大学[3]、熊本大学[4]等の事例が挙げられる。岩手大学では NACSIS 図書コードをシラバスに入力することで連携を行っており、佐賀大学も同様に ISBN 又は書誌 ID をシラバスに入力することで連携している。熊本大学では OPAC の URL を直接リンクしている。このように、コードを利用して連携する方法や、URL を直接リンクする方法等がある。しかしながら、入力者がコードを調べる手間など負担が増えることは極力避ける必要があると考えている。また、既に運用しているシステムの仕様変更を行うことは容易ではない。

そこで、本研究では従来のシラバスシステムの入力部分は仕様変更を行わず連携する手法を開発した。入力されたテキスト・参考書からキーワードを抽出し検索を行う。また、図書館蔵書検索システムの検索機能の仕様変更がされる可能性も十分に考えられる。よって、連携部分も柔軟に対応できるものにしておくことが重要だと考えている。ま

<sup>†1</sup>九州大学大学院統合新領域学府ライブラリーサイエンス専攻  
Department of Library Science, Graduate School of Integrated Frontier  
Sciences, Kyushu University

<sup>†2</sup>九州大学情報基盤研究開発センター学習環境デザイン研究部門  
Learning Spaces Design Section, Research Institute for Information Technology,  
Kyushu University

た本学のシラバスの「テキスト」「参考書」の入力は自由記述のため、そこからキーワードを抽出し蔵書検索システムに渡すことで実現した。

## 2. 連携対象システム

### (1) シラバスシステム

九州大学のシラバスシステムは、これまで部局ごとに別個に整備されていたが、大学内部で開発した全学シラバスへの統合が進行中である[5]。現在 8 学部 8 学府が利用しており、一年間で約 3,500 講義が登録されている。入力欄は 25 項目あり、書籍指定がある場合は「テキスト」「参考書」に入力する。入力形式はテキスト型の自由記述である (図 1)。

授業情報表示	
講義コード	12547213
授業科目名	コアセミナー
講義科目	問題解決学:オペレーションズ・リサーチ入門
授業科目区分	コアセミナー
開講年度	2012
開講学期	前期
曜日時間	木2
の修習状況	
学級	2.0
担当教員	小野 康隆
担当学部署	理工
対象学年	学部1年
開講地区	筑後地区
授業内容	オペレーションズ・リサーチ (OR) とは数学的・統計的モデルにおける「問題」に対して最適な「解」を決定する手法群で、経営学・経済学においてもその手法は広く利用されています。本コアセミナーでは、オペレーションズ・リサーチの入門書を精読することにより、在庫管理、日程計画、ゲーム理論、数値計画といったオペレーションズ・リサーチにおける基本的な問題についてその概要を知ることを目指します。
単位の取得目標	
単位の学習目標	
授業計画	第1回 ガイダンス 第2~15回 テキストの精読。
キーワード	オペレーションズ・リサーチ
授業の進め方	各回、決められた担当教員 (またはグループ) が全員の前で解説をし、また参加者が担当教員に質問をすることをすすめます。
テキスト	松井泰子, 根本俊男, 宇野毅明「入門オペレーションズ・リサーチ」(東海大学出版会)
参考書	
学習環境	本朝日4冊をオフィスアワーとしています。相談がある人は授業41号室まで (無断を避けるため、あらかじめメールにてアポイントメントをとること)、積極的な利用を期待します。
採択/成績評価の方法等	試験/成績評価の方法等 出席30%、報告・議論への参加40%、各回の小レポート30%
その他	教員の個人ホームページ: <a href="http://www.en.kyushu-u.ac.jp/~hrotaka/">http://www.en.kyushu-u.ac.jp/~hrotaka/</a>

図 1 シラバスシステムの出力画面  
 Figure 1 The screenshot of syllabus.

### (2) 図書館蔵書検索システム

本学では、Cute.Catalog という名称の図書館蔵書検索システムが提供されている[6]。機能は、従来の冊子資料を検索する図書館蔵書検索システムに加え、電子ジャーナルや電子ブックなど e リソースのアクセス、機関リポジトリ、デジタル化したコンテンツ等、図書館が提供する多様なコンテンツを集約し、それらの検索機能を提供するものである。オープンソース・ソフトウェア eXtensible Catalog (XC) により構築されたディスカバリ・インターフェースである。Cute.Catalog では、書誌情報を形態素解析や Bi-gram によりインデックス化された情報を元に検索されている。

## 3. システム間の連携手法

### 3.1 システム連携について

図書館蔵書検索システムを利用する際、通常、検索窓に検索語の入力を行うが、今回は、シラバスシステムから検索キーワードを自動で抽出し表示する仕組みである。これにより、情報リテラシーの低い学生でも容易に蔵書検索システムを利用することができる。また、Cute.Catalog の存在を知らなくても、情報探索の導線が確保されるため、書

籍が図書館に蔵書されていることを多くの学生に周知できることが挙げられる。

### 3.2 連携方法

シラバスシステムのプログラムは PHP 言語で記述されている。そのため、今回追加した連携箇所も PHP を用いることにした。連携はシラバスシステムから Cute.Catalog に対して行う。

今回の研究で、仕様変更を行った箇所は主に三点である。一点目は外面的な部分であるアイコンの追加や画面のレイアウトで説明は後述する。二点目は内面的な処理でシラバスの「テキスト」「参考書」からキーワードを抽出するプログラム。三点目も内面的な処理で、シラバスから Cute.Catalog にキーワードを渡すプログラムである。これは Cute.Catalog に渡すための関数を作成し、抽出された各キーワードを引数としている。この仕様変更からも分かるように、利用者が入力する部分はまったく変更していない。

画面レイアウトについては、シラバスの「テキスト」「参考書」欄に「Cute.Catalog で検索」のボタンを追加した (図 2)。検索ボタンをクリックすると入力内容から Cute.Catalog の検索に必要なと思われるキーワードを自動で抽出し、別画面に遷移する (図 3)。赤枠は新たに追加されたフレームである。遷移先画面には抽出した検索キーワードが表示されており、検索ボタンを押下するだけで該当書籍が検索できる (図 4)。また、キーワード毎にチェックボックスを用意しているため、検索に成功しなかった場合は、チェックボックスを手動で操作し、再検索可能である。

テキスト	松井泰子, 根本俊男, 宇野毅明「入門オペレーションズ・リサーチ」(東海大学出版会)
参考書	

↓

テキスト	Cute.Catalog で検索 松井泰子, 根本俊男, 宇野毅明「入門オペレーションズ・リサーチ」(東海大学出版会)
参考書	Cute.Catalog で検索

図 2 テキスト・参考書の欄の変更  
 Figure 2 Changing the fields of textbooks and reference books .

図 2 の実現のために、元のシラバスシステムにおいて、シラバス情報の表示プログラムの二行のみ変更した。`$textbook` と `$refer` はプログラム中で教科書や参考書の情報が記述されている変数であり、`cute` は検索のためのボタンを付加するための関数である。

```
$textbook <input type="checkbox" value="" /> cute($textbook)
$refer <input type="checkbox" value="" /> cute($refer)
```



図 3 シラバスシステムから Cute.Catalog での検索画面  
 Figure 3 The screenshot of search from the syllabus system to the Cute.Catalog system.



図 4 Cute.Catalog の検索結果  
 Figure 4 The screenshot of the result for search in the Cute.Catalog system.

し→④授業中に資料を配布する(適切語)→⑤1文削除(表示なし)

#### 4. テキスト処理によるキーワードの抽出

##### 4.1 形態素解析による判別

シラバスに記載されている教科書と参考書のデータから句を 4,625 個の句を抽出し、人手により適切な句と不適切な句に分類した。適切句は 3,443 個、不適切句は 1,182 個であった。次に、適切句と不適切句の間で句を構成する品詞の種類に際があるのではないかと考え、句を形態素解析した。形態素解析には MeCab[7]を利用した。表 1 に句を構成する品詞の分類結果を示す。

表 1 から、固有名詞は、適切句に多いが不適切句には少ない。動詞は、適切句には少ないが不適切句に少ないということが読み取れる。適切句には著者情報が多く含まれ、また不適切句には書籍に関する説明が多く含まれるからである。しかしながら、「直接九州大学生協伊都ビッグオレンジ書店カウンターで4月末までに支払うこと」のように、大学名やキャンパス名の固有名詞が不適切句に含まれていることがあるため、固有名詞を含む句をすべてキーワード

として選択するのは問題がある。

また、「ジェンダーで学ぶ社会学」は動詞を含むが、適切句であるため、動詞を含む句を削除するのも問題がある。

表 1 句を構成する品詞の分類

Table 1 Classification of parts of speech that make up the phrase.

品詞	適切句	不適切句
名詞	9,590	3,701
固有名詞	(2,745)	(53)
他の名詞	(6,845)	(3,648)
助詞	1,240	2,018
動詞	282	1,100
助動詞	73	461
形容詞	37	45
副詞	52	88
接頭詞	118	49
連体詞	14	34
接続詞	13	49
感動詞	1	0
記号	94	7
合計	81,414	55,408

##### 4.2 サポートベクターマシンによる判別

句を構成する品詞の有無だけでは適切句と不適切句を判別できないため、サポートベクターマシンにより解析した。解析には、統計パッケージ環境である R のライブラリーを利用した[7]。

句を構成する品詞の数から  $x$  を次のように設定した。

$$x = (\text{固有名詞数}, \text{他の名詞数}, \text{助詞数}, \text{動詞数}, \dots)$$

表 2 は SVM による判別結果である。誤判別率は 0.101 であった。

表 2 SVM による判別結果

Table 2 Result of the determination by the SVM.

	不適切句	適切句
不適切句と判別	843	188
適切句と判別	339	3,255

##### 4.3 キーワードの抽出の流れ

以上の考察から、自由記述された内容を元に以下の処理を行った。

- ① アイコンをクリックし、該当する1文を読み込む。
  - ② ①に含まれるデリミタ(句読点, カッコ等. 日本語の場合, 空白も含む)を任意の文字(カンマ)に置換する. カンマを区切りとして取得したものを句とする.
- この句を検索語として扱うが、先述したように自由記述のため、検索語としては不適切な句が含まれている。そこでできる限り不要語を取り除くために以下の処理を行った。
- ③ 検索キーワードとして不適切な単語, たとえば, 編著,

書籍の金額等を削除する。

- ④ 各句をサポートベクターマシン (SVM) にかけて、不要句か否かを判別する。
  - ⑤ 「教科書は指定しない」、「授業中に資料を配布する」のような明らかに不要句と判別できるものは削除する。判別は正規表現で行っている。
- ①～⑤の処理を行うことにより、検索語に適した句をある程度の精度で抽出することができる。

また、図書館蔵書検索システムの検索画面において、表示されている句のチェックボックスは④の処理で不要句と判別されたものは未選択、適切句と判別されたものは選択済みとする。検索時、該当書籍が見つからなければ、チェックボックスを操作し再検索を行うこともできる。

①～⑤の処理の流れを、例文を使用して説明する。

- 例1) ①参考書：政岡光宏編著『初めて学ぶ財務諸表分析改訂版』同文館。→②参考書 政岡光宏編著 初めて学ぶ財務諸表分析 改訂版 同文館→③参考書 政岡光宏 初めて学ぶ財務諸表分析 改訂版 同文館→④参考書 (不要語) 政岡光宏 (適切語) 初めて学ぶ財務諸表分析 (適切語) 改訂版 (適切語) 同文館 (適切語) →⑤参考書 (削除) 政岡光宏 (選択済み表示) 初めて学ぶ財務諸表分析 (選択済み表示) 改訂版 (選択済み表示) 同文館 (選択済み表示)
- 例2) ①一般的な力学のテキストの中から各人に合うものを自分で選んで適宜参照することを推奨する→②変化なし→③変化なし→④一般的な力学のテキストの中から各人に合うものを自分で選んで適宜参照することを推奨する (不要語) →⑤一般的な力学のテキストの中から各人に合うものを自分で選んで適宜参照することを推奨する (未選択表示)
- 例3) ①授業中に資料を配布する。→②授業中に資料を配布する→③変化なし→④授業中に資料を配布する (適切語) →⑤1文削除 (表示なし)

## 5. おわりに

今回は入力内容からキーワードを抽出し、簡単なテキスト処理をしたことで、ある程度の効果は見られたが、性能をさらに向上させる必要がある。書誌情報と書誌に関する説明や資料の有無の文は識別するには、書誌に関する知識や意味解析が必要になる。しかしながら、高度な意味解析をしなくても、本研究で用いた簡単な形態素解析による品詞情報や書誌情報として不適切な語句の抽出、サポートベクターマシン等のような統計的手法によりシステム実装していく予定である。

## 参考文献

- 1) 文部科学省:第3 学生支援・学習環境整備について [http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo4/houkoku/attach/1297027.htm](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/houkoku/attach/1297027.htm) (最終アクセス日 2012/12/20) .
- 2) 遠藤教昭, 高橋里奈, 竹谷隆則, 吉田等明:シラバスデータベースと図書館情報データベースの連携に関する研究, 学術情報処理研究, 10 号, pp.21-28 (2006).
- 3) 佐賀大学附属図書館 図書館メールマガジン「図書館 ML 通信 70 号」2011 年のアーカイブ <http://www.lib.saga-u.ac.jp/public/annual.php?y=2011&m=2&c=nc5&c=ncs1> (最終アクセス日 2012/12/22) .
- 4) 高木貞治:図書館システムと他システムとの連携 熊本大学附属図書館における他システムとの連携事例-図書館システムと大学ポータル統合認証を中心に-, 大学の図書館 27 巻 7 号通巻 No.416 特集, pp.142-146 (2008).
- 5) 九州大学シラバス <http://syllabus.kyushu-u.ac.jp> (最終アクセス日 2012/12/22) .
- 6) 兵藤健志, 工藤絵理子, 越戸陽子, 牧瀬ゆかり, 井川友利子, 大村武史, 片岡真, 星子奈美, 寺田良司:九州大学附属図書館における Cute.Catalog のデザインと開発 - OPAC からディスカバリ・インターフェースへ-, 情報管理, Vol.53, No.6, pp.311-326 (2010).
- 7) MeCab <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> (最終アクセス日 2012/12/22) .
- 8) 豊田秀樹: データマイニング入門-R で学ぶ最新データ解析-, 東京図書 (2008).