

# シラバスデータを利用した関連科目データベースの構築

## Development of Related Subject Database Using Syllabus Data

吉崎 辰悟, 越智 洋司  
 Shingo YOSHIZAKI, Youji OCHI  
 近畿大学理工学部電気電子工学科  
 Faculty of Science and Engineering, Kindai University  
 Email: ochi@kindai.ac.jp

**あらまし**：学生が自分の必要とする講義を探し出す際に、シラバスを利用するという手法が挙げられる。そこで、シラバスの科目ごとに類似性や関連性を把握し関連科目データベースを作成することで、学生が講義の内容をより深く理解するための手助けになるのではないかと考えた。本研究では、近畿大学のシラバスの情報を解析し、文章中の単語の類似度やパターンから関連科目を予測し、それに基づきデータベースを構築する。

**キーワード**：シラバス, 関連科目, 形態素解析, Doc2Vec

### 1. はじめに

大学生が自らの必要とする講義や参考書を探し出す手段として、シラバスを利用する手法が挙げられる。シラバスとは、大学で開講される授業や講義の大まかな内容を示したものであり、講義の目的や各時限の授業内容などが記載されている。近年では、シラバス検索システムを導入する大学も一般的になっており、自分の受講する講義の情報を知ることが容易となった。我々は、科目ごとの類似性や関連性を把握し、関連科目データベースを作成することで、学生が講義の内容をより深く理解するための手助けになるのではないかと考えた。

本研究では、シラバス本文中の単語を解析し、その類似度や頻度から関連科目を予測可能なデータベースを構築する。

### 2. 先行研究

本研究の参考となる研究として、シラバスに関する先行研究について述べる。川場<sup>(1)</sup>らは、ウェブシラバスを公開している高等教育機関のシラバスを収集・分析し、汎用型ウェブシラバスシステムを開発している。また、堀ら<sup>(2)</sup>は、シラバスの情報を利用することで履修決定支援や個人のカリキュラムの特徴を見出す研究を行っている。

### 3. システム概要

作成した Web アプリケーションの概要について述べる。初めにユーザは、インタフェースで検索内容の入力を行う。次に、その検索内容からデータベースより条件に合ったデータを受け取り、ユーザに開示する。表示された科目を選択すると、関連科目の分析が開始されユーザに開示する。本研究では、データベースの管理として Amazon SimpleDB を用いる。本システムの構成を図 1 に示す。

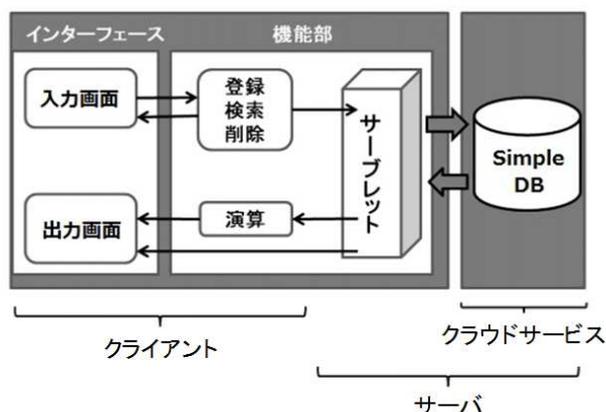


図 1 システム構成

### 3.1 シラバスデータベースの構築

本研究では、近畿大学理工学部電気電子工学科の189科目分のデータを対象に収録した。データベースへの登録内容としては、科目名や教員名、科目区分、開講期などである。

シラバスデータテーブルの登録手順について述べる。まず、近畿大学のシラバスサイトにアクセスし、Web ページよりシラバス HTML ファイルを取得する。次に、HTML ファイルを HTML Parser を用いて解析する。本研究では、HTML Parser として「NekoHTML<sup>(3)</sup>」を使用する。NekoHTML によりシラバスから科目名や教員名をはじめとする情報を取得する。最後に、取得したデータを SimpleDB に格納する。図 2 にデータ登録の流れを示す。

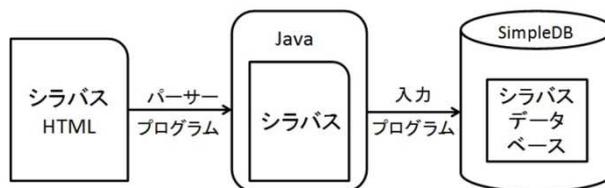


図 2 データ登録の流れ

### 3.2 関連科目の導出

関連科目の導出方法について述べる。関連性を見出す方法として、文書内部の単語を比較しベクトルを割り当てることで類似度を導出する手法がある。本研究では、シラバス本文中の単語を形態素解析し、その類似性から関連科目を抽出する。なお、類似性の分析については Doc2Vec を使用した。以下、実装の方法について述べる。

#### (1) 形態素解析

形態素解析とは、文章をある単語に区切り、辞書を利用して品詞や内容を判別することである。本研究では、形態素解析器として「kuromoji<sup>(4)</sup>」を使用した。シラバスの文章から名詞のみを抽出し、メタデータの作成を行う。

#### (2) Doc2Vec

Doc2Vec とは、単語をベクトル化して定量化する手法に基づいて、文書の類似度やベクトル計算などを実現する手法である<sup>(5)</sup>。通常、1 文書ごとに単語を羅列して文書間を比較するのに用いられるが、本研究では1 科目ごとの単語集合を1 文書分として科目間の類似性を抽出する。本研究では、実装ライブラリとして chainer<sup>(6)</sup>を利用した。

### 3.3 科目検索システムの開発

ユーザが容易に目的となる講義を探し出せるよう、科目検索システムを作成する。シラバスデータベースの中から、ユーザの指定したキーワードが含まれる科目名のリストを表示する。また、様々な条件で絞り込みができる詳細検索画面の作成も行う。

### 3.4 ユーザインタフェース

UI の作成として「GWT Bootstrap」を用いる。GWT Bootstrap はレスポンシブル Web デザインに対応しており、ブラウザの横幅サイズを判断基準としてレイアウトデザインを柔軟に調節する事が可能である。

本研究では、検索画面、科目一覧画面、関連科目検索結果画面を作成。科目一覧画面で科目を指定すると、Doc2Vec により導出した関連科目が表示される。図3に関連科目検索結果の実行画面を示す。

科目名	学部	学科	教員名	科目分類	年次
エネルギー伝送工学_工ネ	理工学部	電気電子工学科	武村祐一郎	専門科目	3年次
ネットワーク工学_工し機	理工学部	電気電子工学科	越智洋司	専門科目	3年次
電子デバイス_情報-総合	理工学部	電気電子工学科	前田佳伸	専門科目	4年次
科学的問題解決_電	理工学部	電気電子工学科	小川壽弘	共通教養科目	3年次
電磁気学3_工し機_工ネ	理工学部	電気電子工学科	松本芳幸, 森本健志	専門科目	3年次

図3 実行画面

### 4. 評価実験

導出した関連科目（以下、類似度関連科目）の妥当性を検討するために、シラバスに明記されている関連科目（以下、シラバス関連科目）と類似度関連科目上位5件が対応しているかを確認する方法<sup>(7)</sup>で行う。評価には式(1)の適合率と式(2)の再現率を用いる。なお、式中の変数は表1の混合行列に対応している。

近畿大学理工学部電気電子工学科の専門科目について検証を行った結果、適合率は34.7%、再現率は24.3%となった。

$$\text{適合率} = \frac{TP}{TP + FP} \quad \text{式(1)}$$

$$\text{再現率} = \frac{TP}{TP + FN} \quad \text{式(2)}$$

表1 混合行列

		類似度関連科目	
		あり	なし
シラバス 関連科目	あり	TP = 66	FN = 206
	なし	FP = 124	TN

### 5. おわりに

本稿では、関連科目検索システムのためのデータベースと検索システムの開発について述べた。本評価実験においては、「類似度関連科目には存在するが、シラバスには関連科目として記載されていない科目 (FP 科目)」の妥当性については判断していない。シラバス関連科目は教員の主観が入ったものであり、FP 科目の中には、関連性が高いにも関わらずシラバスに関連科目として記載されていない科目も存在すると考えられる。今後、FP 科目に関する教員の主観評価をする予定である。また、本実験では類似度に関わらず上位5件を対象としたため、精度を上げる閾値についても検討をする。

#### 参考文献

- (1) 川場 隆, 土屋 健, 小柳 恵一: “汎用型ウェブシラバスシステムの開発”, 日本教育工学会論文誌 35(Suppl.), pp61-64, 2011
- (2) 堀 幸雄, 中山 堯, 今井 慈郎: “カリキュラムの特徴抽出と時間割の要約生成”, 情報知識会誌, Vol.20, No.2, pp201-206, 2010
- (3) NekoHTML - CyberNeko HTML Parser, <http://nekohtml.sourceforge.net/>, accessed Nov.7, 2016
- (4) kuromoji - japanese morphological analyzer, <http://www.atilika.org/>, accessed July.19, 2016
- (5) Quoc Le, Tomas Mikolov: “Distributed Representations of Sentences and Documents Proc”, Of ICML2014, pp1188-1196, 2014
- (6) Chainer: A flexible framework for neural networks, <http://chainer.org/>, accessed Nov.14, 2016
- (7) 株式会社システム計画研究所: “Python による機械学習入門”, 第1版, オーム社, pp50-51(2016)