

# 県内大学シラバスリポジトリを用いた関連授業検索システムの開発

2004MT045 北折 佑介 2004MT112 梅田 修

指導教員 河野 浩之

## 1. はじめに

各大学でシラバスシステムの構築が進むにつれ、リポジトリの構築や大学間のシラバスの内容やメタデータの比較を行う研究がされている。しかしこれらの研究を利用した実用的なシラバスシステムの機能はあまり実装されていない。

本研究では愛知県単位互換事業にこれらの研究を応用し、南山大学瀬戸キャンパスの授業と類似している互換事業提携大学で受講可能な授業を表示する「他大学関連検索機能」をCMS上に実装する。そのために、シラバスに記述されている授業名と授業内容を用いクラス分類を行う。最後に、分類結果を関連検索機能として実験的に CEAS<sup>\*1</sup> に実装する。

## 2. 教育用 CMS の選択と WEKA

### 2.1. シラバスに関する先行研究[1][2]

シラバスに関連した先行研究は、主に以下の 3 つについて行われている。

- リポジトリの構築
- メタデータの統一
- 文章内容比較

その中で、Manas Tungare ら[2]はオープンソースの CMS である Moodle を用いてシラバス作成機能を追加。教員がシラバスを作成するときに、授業内容に関連性のある他大学のシラバスを参照可能にし、教員をサポートする機能を提案している。この研究では、教員と生徒に類似した講義を提案できることを目標としており、リポジトリ拡大のためにシラバスの自動収集が行われている。

### 2.2. 教育用 CMS の比較

我々は多数にある CMS の中から用途、目的に合った CMS を選択する必要がある。そこで我々は先行研究でも挙げられていた CMS も含め、計 6 つの CMS のライセンスと特徴の比較を表 1 で記す。

表 1 CMS 比較表[3]

	ライセンス	特徴
CEAS 2.2.0	制限有	シラバス作成の容易さ、レイアウトが綺麗で分かりやすい
Moodle 1.9	GPL	世界的ユーザの多さ、サポート、拡張の充実度が他 LMS と比べ格段に高い

Plone 1.0	GPL	ワークフロー、権限設定等細かな設定が出来るがアプリ、プログラム、DB 言語が特別に必要な知識が多い
exCampus 2.0.0	制限有	ストリーミングサーバを準備すれば動画配信が可能
Site@school 2.4.10	GPL	海外で幼稚園の学習にも利用出来るほど単純で判りやすい
Sakai 2.4.1	ECL ver.1.0	PDA 等の携帯電話端末表示対応機能

我々は CMS を比較した結果、シラバスリポジトリ作成による関連検索機能が実装できるだけでなく、シラバスの作成、編集が簡単であり、レイアウトが綺麗で分かりやすくサイト全体の管理もしやすい CEAS を選択する事にした。なお、CEAS のライセンスは限定的な条件下での使用を許可するものであり、再配布や商用目的での利用は一切許可をしていない。

## 3. シラバスシステムの構築

### 3.1. リポジトリの構築

シラバスのクラス分類を行う際に、まず関連させたい大学のシラバス(実験データ)を集める必要がある。シラバスの収集後、クラス分類のプログラムに用いるためファイル形式を txt 形式の文章ファイルに統一を行う。

### 3.2. クラス分類の全体の流れ

クラス分類までの全体の流れが下図 1 である。

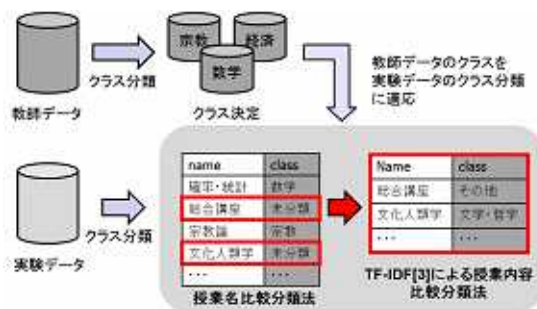


図 1 分類までの流れ

\*1 CEAS Community Page, <http://ceascom.iecs.kansai-u.ac.jp/>

初めに、実験データのクラス分類の基準となる教師データのクラス分類を行う。次に、シラバスの授業名を用いた比較を行う「授業名比較分類法」でクラス分類する。そして、「授業名比較分類法」で分類できなかったものは、授業内容で分類する「TF-IDFを用いた授業内容比較分類法」を用いて分類をしていく。

全てのシラバスを授業内容だけで分類しない理由は、大学ごとのシラバスの形態が様々であり、授業内容を機械で分類すると必要な情報が記載されていないシラバスや、様々な単語が文書内で出現するシラバスを正確に振り分けの事が非常に困難になる為である。

### 3.3. 教師データのクラス決定

他大学のクラス分類を行うために教師データとなるシラバスのクラス分類をしなくてはならない。クラス分類の方法は、各大学で決められている学科や学部を参考に分類を行っていく。教師データのクラス決定は後のクラス分類のルールにも影響を及ぼす為、ある程度結果を想定した上での決定が望ましい。その為、どのクラスにも属さないシラバスは「その他」クラスと定め、参照の対象としないようにする措置が必要である。

### 3.4. 授業名比較分類法

シラバスのクラス分類は、初めに各シラバスの授業名を用いた「授業名比較分類法」を用いて行う。この方法は、シラバスの授業名には一番重要な単語が記述されており、文章内容よりも明確にクラスが判別できることに注目したものである。

まず初めに、形態素解析ツール茶釜(ver.2.1)<sup>\*2</sup>で授業名から名詞のみを抽出する。実験データからは各授業名の単語を抜き出し、あらかじめクラス分類されている教師データからはクラスごとに比較単語リストを生成する。そして比較単語リストと実験データから抽出した単語を比較し、実験データにクラスを割り当てる。

### 3.5. TF-IDFを用いた授業内容比較分類法

「授業名比較分類法」によりクラス分類を行った後、分類できなかったデータを「TF-IDFを用いた授業内容比較分類法」により、シラバスの文章を使った分類を行っていく。本研究では、Perl モジュール「TermExtract<sup>\*3</sup>」を用いて専門用語の抽出と文章の重み付けを行う。

まず授業内容で分類を行うために、教師データから重要単語の抽出を行い、重要単語リストを作成する。次に、重要単語リストを用いて教師データと実験データを TF-IDF で重み付けする。TF-IDF とは、一つの文章中で特定の単語の出現率が高く、全文書中でのその単語出現率が低いほど重みが高くなる手法である。TF-IDF は図 2 の式で表すことができる。そして、クラス分類はデータマイニングツール

WEKA(ver.3.4.11)<sup>\*4</sup>による決定木アルゴリズム J48 を用いて教師あり分類する。WEKA は、Waikato 大学が中心となって開発しているツールである。

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$w$  : TF - IDF の値  
 $i$  : 目的単語  $j$  : 文書  
 $tf$  : 対象ドキュメント内の目的単語の出現数  
 $N$  : 比較する全ドキュメント数  
 $df$  : 目的単語が出現したドキュメントの総数

図 2 TF-IDF

## 4. 南山と他大学シラバスのクラス分類

### 4.1. リポジトリの構築

南山大学瀬戸キャンパスと「愛知学長懇話会・単位互換事業」で定められている愛知県の四年制大学を対象にシラバスを収集する。それに加え南山大学と連携聴講生制度で提携している豊田工業大学と南山短期大学のシラバスも収集を行う。対象とするのは 2007 年度秋学期のシラバスであるが、南山大学のみ 2007 年度春学期も対象とした。それらの結果、収集対象となった南山大学瀬戸キャンパスの全シラバス 821 個のうち 574 個の収集をした。

単位互換制度に提携している大学中、2007 年度秋学期に南山大学で受講可能な教科は豊田工業大学 86 個、南山短期大学 12 個、その他の大学 103 個で合計 201 個ある。多くの大学が数教科の授業しか公開しておらず、最高で愛知県立大学の 17 個が最大であった。そのためツールによる自動収集は効率が悪いと予想されるため、検索エンジンにより手動で収集を行った。その結果、計 185 個の他大学のシラバスを収集する事に成功した。

### 4.2. 教師データのクラス決定

教師データの作成のために南山大学のシラバス全ての授業にクラスを割り当てていく。最終的にこれらのクラスを他大学のシラバスに割り振ることになる。教師データのクラス割り当てをする為に我々は南山大学瀬戸キャンパスの全授業を 20 のクラスに細分化した。表 2 はそのクラス一覧である。

<sup>\*2</sup>ChaSen's Wiki, <http://chasen.naist.jp/hiki/ChaSen/>

<sup>\*3</sup>専門用語自動抽出用 Perl モジュール Term Extract の解説, <http://genshen.dl.itc.u-tokyo.ac.jp/termextract.html>

<sup>\*4</sup>Weka 3 - Data Mining with Open Source Machine Learning Software in Java, <http://www.cs.waikato.ac.nz/ml/weka/>

表2 分類クラス一覧

プログラミング・ソフトウェア	10	文学・哲学	5
ネットワーク	9	日本語	20
情報その他	26	英語	138
OR・解析・統計	16	外国語その他	46
数学	21	宗教	19
物理・化学・生物	8	心理	4
工学	10	教育	27
環境	21	音楽・体育・美術	59
経済	23	その他	18
歴史・社会・文明・国際	48		
法律・倫理・政治	46	合計	574

4.3. 授業名比較分類法による分類

教師データの分類後、南山大学瀬戸キャンパスの教師データを用い、他大学のシラバスをクラス分類する。下図3は「授業名比較プログラム」の実行結果として作成された「class.csv」を開いたものである。

1	教科名	A	B	C	E
2	中部大学 公衆衛生学.txt	[衛生2]	[公衆1]	環境	
3	中部大学 経営ケース分析.txt	[経営6]	[分析2]	経済	
4	中部大学 総合芸術論.txt	[芸術2]			
5	中部大学 英蘭の思想と表現.txt	[英蘭2]	体育・美術・音楽	[表現1]	語
6	中部大学 科学史.txt	[科学1]	志望・倫理・政治	[科学2]	理
7	中部大学 電気電子物性A.txt				
8	中部大学 音楽の表現と歴史.txt	[音楽1]	体育・美術・音楽	[表現1]	語
9	南山短期大学 コミュニケーション論B.txt	[CoRa-コミュニケーション44]	英語		
10	南山短期大学 スピーチ批評B.txt				
11	南山短期大学 多文化共生論B.txt	[共生1]	[文化9]	歴史・社会・文明・国際	
12	南山短期大学 第三世界の女性と子供.txt				
13	南山短期大学 英語の語法B.txt	[英語108]	英語		
14	名古屋理科大学 加藤立国としての日本.txt	[日本1]	歴史・社会・文明・国際	[日本1]	法
15	名古屋大学 ドイツ文学講義.txt	[文学2]	文学・哲学		
16	名古屋大学 人間発達科学1.txt	[人間10]	法律・倫理・政治	[科学2]	理
17	名古屋大学 技術教育学講義.txt	[技術2]	[教育18]	教育	
18	名古屋大学 政治学.txt	[政治12]	法律・倫理・政治		
19	名古屋大学 教育と発達心理学.txt	[教育18]	[心理1]	教育	
20	名古屋大学 現代社会と教育.txt	[教育18]	[現代1] 9:	教育	
21	名古屋大学 現代社会と志望.txt	[現代1]	[社会1] 5:	法律・倫理・政治	
22	名古屋大学 科学・技術の倫理.txt	[技術1]	[倫理1]	志望・倫理・政治	[科学2]
23	名古屋女子大学 身体表現I.txt	[表現1]	情報その他	[表現2]	理
24	名古屋工業大学 保存修復学.txt	[表現1]	情報その他	[表現1]	目
25	名古屋工業大学 外国語(洋英)イテリ.txt	[E-イテリ6]	[外国10]	外国語その他	
26	名古屋工業大学 技術と文明.txt	[技術1]	日本語	[文明15]	歴
27	名古屋工業大学 構造デザイン学.txt	[構造1]	歴史・社会・文明・国際		

図3 授業名比較法分類結果

Class は3つまで出力され、何も出力されなかったものは合致する単語が無かった未分類のシラバスである。分類結果が表示されたものは134個で、空欄になっている授業は51個あった。この中には「その他」も含まれており、「その他」は未分類として空欄で出力されている。

4.4. TF-IDFを用いた授業内容比較分類法による分類

「授業名比較分類法」により分類結果が出力されなかった授業を、第二段階の分類である「TF-IDFを用いた授業内容比較分類法」で分類した。これらは「授業名比較分類法」により出力されなかった51個のシラバスを対象とし、図4がその結果である。

	A	B	C
中部大学 総合芸術論		英語	
中部大学 電気電子物性A		その他	
南山短期大学 スピーチ批評B		その他	
南山短期大学 第三世界の女性と子供		その他	
名古屋工業大学 保存修復学		その他	
名古屋工業大学 テーマ科目(インペーション)とは何か		その他	
名古屋工業大学 微生物薬品学I		その他	
名古屋工業大学 生産学講義		その他	
名古屋工業大学 機能性食品学II		経済	
名古屋工業大学 食品安全学		法律・倫理・政治	
名古屋工業大学 食生活論		その他	
愛知大学 消費者保護法		その他	
愛知大学 総合科目(共通)		宗教	
愛知学院大学 総合科目III-II		心理	
愛知学院大学 総合科目II-II		教育	
愛知県立大学 国語概論I-II		その他	
愛知県立大学 研究各論(イギリスの歴史A)		法律・倫理・政治	
愛知県立大学 障害者福祉論		その他	
東海学園大学 医学一般I		その他	
福山女学院 シェンダー・セクシュアリティ論II		その他	
豊橋創造大学 総合講座		歴史・社会・文明・国際	
豊田工業大学 パワーエレクトロニクス		日本語	
豊田工業大学 メカトロニクス		日本語	
豊田工業大学 力学1および演習		OR・解析・統計	
豊田工業大学 力学2		その他	
豊田工業大学 力学3		その他	

図4 授業内容比較法分類結果

4.5. 分類精度の算出

本研究の他大学の分類精度は、「授業名比較分類法」と「TF-IDFを用いた授業内容比較法」二つの分類法それぞれの精度と、二つの手法を合わせた場合の精度の三つを算出する。結果が表3である。

表3 分類精度の精度表

手法	対象シラバス数	正解シラバス数	分類精度
授業名比較分類法	185	109	59%
TF-IDFを用いた授業内容比較分類法	51	12	23%
授業名 授業内容	185	121	64%

この結果を見てみると、授業名比較法での精度は6割という結果となったが、残りシラバスの授業内容比較法の精度は2割ほどしかなかった。これでは南山大学瀬戸キャンパスにない未知の授業が出てきた際の分類法としての精度は不十分であり、総合結果としても実用レベルには至っていない。

4.6. TF-IDFを用いた授業内容比較分類法の改善と精度評価

授業名分類法で結果が出力されなかった51個のうち17個が豊田工業大学のシラバスであった。分類できなかった豊田工業大学のシラバスのほとんどは、「物理・化学・生物」クラスであったが、南山大学にはこれらの教師となるような授業データが圧倒的に不足していた。ゆえに、教師データのシラバスの数がクラスによって不足していることが正確に分類されなかった原因だと考えた。よって、この問題を解決するため、教師データでシラバスの数が不足しているクラスに新たにシラバスを追加した。

追加したシラバスは、名古屋大学 2007年度シラバスの工学部、物理系授業の65個であり、これらを教師データに追加した。そして教師データ変更後の他大学を「TF-IDFを用いた授業内容比較分類法」で分類した結果、図5のよ

うに授業内容で新たに 16 個のシラバスを正しく分類し、精度の向上が見られた。

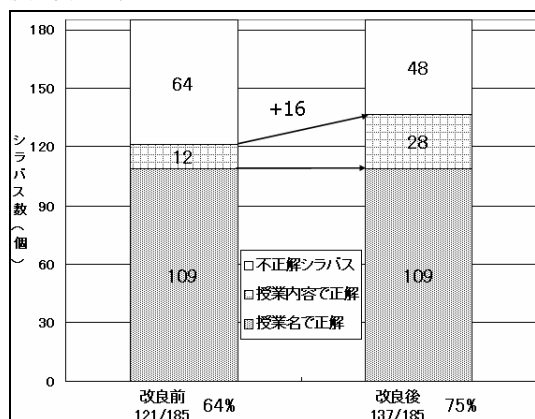


図5 改良後の精度比較グラフ

## 5. 他大学関連検索機能の実装

### 5.1. CEAS のインストール

我々は CEAS のインストールを図 6 の動作環境で行った。

■ PC : PC STATION PT875
■ プロセッサ : AMD Athlon 64 X2 4000 +
■ メモリ : 1GB
■ OS : Linux Fedora 7

図6 CEAS 動作環境

### 5.2. 他大学関連検索機の実装

4 章でクラス分類した他大学のシラバスを実際に CEAS に登録し関連検索機能を実装する。CEAS で使用されるデータベース (CEASDB) に、新たに南山大学の授業名、クラスが書かれたテーブルと他大学の大学名、授業名、クラス、URL が書かれたテーブルの二つを追加する。

その後、"www/html/System/student\_men" 内にある class\_top.php の 333 行目に、他大学授業のリンクを表示するためのプログラムを追記した。このプログラムは、まず南山のシラバスを開いたとき、そのシラバスの授業名を取得する。次に、その授業名のクラスを南山のテーブルから読み込む。読み込んだクラスと他大学のテーブルに登録されたクラスを比較し、同じクラスが他大学の中にあるか探す。そして、一致した項目全てに対し、テーブル内にデータベースから大学名と授業名を読み込み表示し、授業名にはシラバスへのリンクを貼る。

今回は実験として、CEAS 内に他大学のシラバスを登録し、そこへリンクすることにした。

図 7 は実際に CEAS に他大学関連検索機能を実装した

部分のスクリーンショットである。

### 授業実施画面

科名	確率・統計
担任者名	高橋 俊也
系・学科・コース	金学科
年度	2007
学期	春
曜日	
時間	

他大学関連授業	
大学名	授業名
名古屋市立大学	統計学2
豊田工業大学	ベクトル解析
豊田工業大学	数値解析
豊田工業大学	数値計算法

図7 関連検索機能実装図

## 6. まとめ

本研究では、他大学関連検索機能の実装をするために、授業名と授業内容の二つを用いて、シラバスの比較と分類の方法を提案した。この結果、教師データと実験データのシラバスの分野傾向や数、そして大学ごとのシラバスの記述の形態によっても精度にばらつきが出る事が分かった。本研究では教師データに足りないシラバスを補うことでこの問題を解決し、分類の精度を 75% に上げる事に成功した。

本研究の最後に、分類結果を用い実験的に CEAS に他大学関連検索機能を実装した。

## 参考文献

- [1] 山田信太郎, 伊東栄典, 廣川佐千男, "Web 上に公開されたシラバス情報の自動収集," 九州大学大学院システム情報科学府 九州大学情報基盤センター, <http://rd.cc.kyushu-u.ac.jp/~itou/2002/dicom2002.pdf>, pp.1-4, 2002, (accessed 2007.8).
- [2] Manas Tungare, Xiaoyan Yu, William Cameron, GuoFang Teng, Manuel A. Perez-Quinones, Lillian Cassel, Weiguo Fan, Edward A. Fox, "Towards a Syllabus Repository for Computer Science Courses," Virginia Tech and Villanova University, [http://manas.tungare.name/publications/tungare\\_2007\\_towards.pdf](http://manas.tungare.name/publications/tungare_2007_towards.pdf), pp.1-5, 2007, (accessed 2007.8).
- [3] 田中裕也, 井ノ上憲司, 市川尚, 鈴木克明, "教育用 CMS の比較分析と選択支援サイト OCET の開発," 教育システム情報学会研究報告集 vol.20-1, <http://www2.et.soft.iwate-pu.ac.jp/~g231d019/files/ocet.pdf>, pp. 43-48, 2005. (accessed 2007.8).