

# 編入学生単位認定の作業効率化のためのシラバス比較ツール

三好 康夫

高知大学 理工学部情報科学科

miyoshi@is.kochi-u.ac.jp

## A Syllabus Comparison Tool to Support Credit Recognition Tasks for Transfer Students

Yasuo MIYOSHI

Dept. of Information Science, Fac. of Science and Technology, Kochi Univ.

### 概要

多くの大学で、入学生が入学前に他大学・高専等で履修した単位を入学後の大学で修得した単位として認定する規則が整備されている。この制度は高専等からの編入学生が4年次で卒業するためには必須のものであるが、どの科目の単位を認定して良いかの判断は、当該科目の担当教員やその分野の専門知識を有する教員でなければ困難である。本研究では、国立高専のシラバスが全て統一されていることから、国立高専シラバスと大学シラバスの比較ツールを開発し、国立高専からの編入学生の単位認定が可能な科目を調べる作業の効率化を行った。

### 1 はじめに

著者の勤務先の学科では、大学院進学者を増やしたいという目的で、高専からの3年次編入生の受け入れを大幅に増やすことにした。大学には入学生が入学前に他大学・高専等で履修した単位を入学後の大学で修得した単位として認定する規則が整備されており、編入学生にとっては、4年次で卒業するためには必須の制度である。現在の学科の方針としては、できる限り学習済みの科目については単位を認め、大学院科目の早期履修制度の活用を促すことで、大学院進学希望者を増やしたいという考えである。学生は、早期履修により修士課程進学後、研究に専念する十分な時間が確保できる。

しかし、単位認定の申請やその可否の判定は、入学後かなり短い期間で行う必要があり、また詳細は割愛するが教員には非常に煩雑な作業が求められる。特に、学生に単位認定願の書類を作成させる際に、学生にはその学生のアドバイザー教員と十分に話し合った上で単位の認定を申請する科目を決めさせる必要がある。これは情報科学分野特有の問題かもしれないが、既修得科目と本学の科目とが1対1で対応できないことが多い。したがって、複数科目の組み合わせが必要で、どう組み合わせると認定単位数を多くできるかという組み

合わせのパターンまで考慮すると非常に難しくなるからである。

どの科目の単位を認定して良いかの判断は、既修得科目と本学の科目のシラバスを見比べて、内容が一致しているかを確認した上で行う。そのため、申請されたものの判定だけであれば、それぞれの科目のシラバスを開いて見比べるだけで良いが、単位認定できそうな科目の組み合わせを探すとすると、全ての修得済科目のシラバスを隅々まで確認しなければならず、非常に重労働である。そこで、このような作業の効率化を目的としたシラバス比較ツールの開発を行った。本稿では、開発したツールの概要と実際に利用して得られた知見について報告する。

### 2 シラバス比較ツールの概要

国立高専では、全国51高専のシラバスが統一したフォーマットで作成され、同じサイト[1]上で公開されている。そのため、国立高専シラバスと著者の勤務先の学科のシラバスをクロールし、シラバスに含まれる専門用語から科目間の類似度を推定し、内容の近い科目を見つけやすくするシラバス比較ツールを開発した。

#### 2.1 シラバスのクロール

国立高専のシラバスのクロール機能は、

Python で Requests と BeautifulSoup 4 のライブラリを用いて実装した。著者の勤務大学のシラバスは非常に収集しにくい形式で公開されていたため、さらに Selenium ライブラリも用いて Web ブラウザの自動操作によるクローリングが必要であった。

高専シラバスは、URL のクエリに school\_id, department\_id, year を指定すると、school\_id で指定した高専の department\_id で指定した学科やコースで、year で指定した年度に開講される科目の一覧が表示される。さらに、科目コードの subject\_code を指定すると、その科目のシラバスが表示されるが、この時の year の指定には注意が必要である。例えば、2024 年度に高専を卒業した学生（2020 年度高専入学）が 4 年生の時（2023 年度）に単位修得した科目のシラバスを見るには、その科目の対象学年が 4 年であれば、year には入学年度の 2020 を指定する。これにより、この学生がその科目の単位を修得した 2023 年度のシラバスを取得できる。例えば、2025 年度開講科目一覧から各科目のリンクを辿っても良いが、この時、対象学年が 1 年の科目へのリンクであれば year=2025 の 2025 年度入学生対象の科目のシラバスが開き、対象学年が 2 年の科目であれば year=2024 の 2024 年度入学生（つまり、2025 年度に 2 年生の学生）対象の科目シラバスが開くことに注意が必要である。

現在は、個々の編入学生の単位修得科目に合わせて高専シラバスのクローリングを行っているのではなく、編入学生の出身の高専・学科の school\_id, department\_id と高専入学年度の year を指定して、開講されている全ての科目のシラバスを収集している。

## 2.2 科目間の類似度推定

科目間の内容が似ているほど、シラバスに含まれる専門用語が多く一致するはずであるという簡易な仮説を立て、科目間の類似度推定を行った。

高専シラバスからは「到達目標」「教育方法等」「授業計画」等の項目に記入されたテキストデータから専門用語を抜き出した。大学の学科専門科目のシラバスからも同様に、授業内容を記した項目より専門用語を抜き出した。テキストデータから専門用語を抜き出す手順を以下に示す。

- (1) 表記揺れをなるべく揃えるため、英数文字は半角に、それ以外は全角に統一する。
- (2) 記号等は半角スペースに置換した上で、MeCab (IPAdic) を使って形態素解析を行う。
- (3) pytermextract モジュール[2]を用い、複合語を

キー、複合語の出現回数を値としたディクショナリを得る。

- (4) 専門用語として含めるべきではない語（図 1）をストップワードとして除外（現在は大学シラバスのみ）し、複合語の出現回数順にソートしたリスト（図 2, freq\_list）を得る。

```
stop_words = [  
    '手法', '意味', '内容', '理解', '方法',  
    '授業', '演習', '講義', '学習', '習得',  
    '習う', '学ぶ', '学び', '関係', '具体例',  
    '科目', '基礎', '概要', '概説', '概論',  
    '目的', '例', '中間試験', '期末試験', '定期試験',  
    '試験', '性質', '進め方', '応用', '講義形式',  
    '紹介', '定義', '値', '概念', 'まとめ',  
    '実例', '業務', '講師', '成果', '知識',  
    'レポート', '関連事項', '説明', '記述', '注意',  
    '基本', '法則', 'ガイダンス', '枠組み', '最初',  
    '使い方', '一つ', '特徴', '身', '範囲', '提出',  
    '通り', '解説', '到達目標', '一連', '考え方',  
    '基礎知識', '基', '基本法則', '前半', '後半',  
    '課題', '学期', '自分', 'グループワーク', '復習',  
    'つぎ', '等', '種々', '目標', '分野',  
    '見方', '知見', '意義', '通常', '応用例',  
    '原因', '実習', '外部', '働き', '既存',  
    '考察', '予備', '前者', '後者', '展望',  
    '側面', '授業実施方法', '演習形式', '練習問題',  
    '仕組み', '他', 'もと', '視点', '原理',  
    '理論', '授業内容', '必要性', 'イントロダクション',  
    '相互評価', '成果発表', '書き方', '基本知識',  
    '確認テスト', '中心', '流れ', 'ポイント'  
]
```

図 1 シラバスに多く含まれるが科目間の類似度推定には不要なストップワード

```
[["ネットワークセキュリティ", 9], ["OSI 参照モデル", 7], ["役割", 7], ["通信", 6], ["インターネット", 5], ["TCP", 5], ["IP", 5], ["プロトコル", 5], ["脆弱性", 5], ["Web アプリケーション", 4], ["IP アドレス", 4], ["トランスポート層", 4], ["コンピュータネットワーク", 3], ["暗号技術", 3], ["物理層", 3], ["データリンク層", 3], ["対策", 3], ["技術", 2], ["暗号化技術", 2], ["特性", 2], ["攻撃手法", 2], ["各層", 2], ["セキュリティ", 2], ["イーサネットフレーム", 2], ["コリジョン", 2], ...]
```

図 2 pytermextract で抽出した専門用語の出現回数順リスト(freq\_list)の例

現在の実装では、pytermextract で抽出した複合語から図 1 のストップワードを除外したものを専門用語としている。termextract には語の出現頻度等から複合語の重要度を計算する機能を有してい

るようで、FLR 法[3]等を用いた専門用語の抽出もできるかもしれないが、図 1 のストップワードを除外するだけで十分と判断した。

高専の科目と大学の科目のそれぞれの freq\_list から科目間の類似度を表すスコアを算出する。この時、大学の科目のシラバスにおいては、科目のシラバス全体の freq\_list だけでなく、各回の授業計画の項目を元に、前半・中盤・後半（全 15 回の授業科目であれば、前半=1~5 回、中盤 6~10 回、後半 11~15 回）のそれぞれでも freq\_list を生成しておく。高専科目と大学科目のそれぞれの freq\_list を比較し、専門用語が一致した総回数を  $m_{count}$ 、一致した専門用語の数を  $m_{words}$  として、一致度  $M$  を次式で計算する。

$$M = \log_2(m_{count}) \times m_{words}$$

一致度は、全体  $M_{all}$ ・前半  $M_{first}$ ・中盤  $M_{middle}$ ・後半  $M_{final}$  それぞれで計算し、最終的な科目間類似度の  $Score$  はこれらの総和で算出する。

$$Score = M_{all} + M_{first} + M_{middle} + M_{final}$$

したがって、科目間で専門用語が一致していた場合、それが大学の授業科目のシラバスの各回の授業計画に含まれていると、より  $Score$  は高くなる。

### 2.3 ユーザインタフェース

高専・学科ごとに科目間の類似度推定の結果を 1 つの HTML ファイルに出力しており、この HTML ファイルを Web ブラウザで開くと図 3 のようなシ

ラバス比較ページが表示される。

シラバス比較ページには、大学の学科専門科目のリストが青文字で表示され、開閉が可能となっている。大学の科目を開くと、その科目のシラバスに含まれている専門用語の一覧が表示される。これは freq\_list を元に表示しており、出現頻度の高い順に並んでいる。そして、類似する高専の授業科目がある場合は、スコアの高い順に緑文字で表示される。デフォルトではスコアが 50 以上の高専科目のみが表示されるようになっているが、「スコア 10~50 の高専科目も表示する」や「スコア 10 未満の高専科目も表示する」というチェックボックスを用意しており、これにチェックを入れると類似度が低いと推定された科目も表示できる。図 3 はスコアが 10~50 の科目も表示するように設定された例である。科目名の右にはそれぞれ、その科目の Web シラバスへのリンクボタンと概要ボタンがある。概要ボタンをクリックすると、ページ右の領域に、クローリングで収集したシラバステキストを用いて科目の概要が表示される。

緑文字で表示された高専の科目名も開閉が可能で、開くと大学の科目と同様に専門用語の一覧が表示される。この時、大学の科目と共通の専門科目は黄色くマークされ、どのようなところが類似しているのかが分かりやすくなっている。また、全体  $M_{all}$ ・前半  $M_{first}$ ・中盤  $M_{middle}$ ・後半  $M_{final}$  の

図 3 シラバス比較ページ

一致度をそれぞれ表示し、どこでどの専門用語が一致しているかがわかるようにしている。これにより、仮にスコアが高い高専科目であっても、大学の授業科目の前半で扱っている内容しか含まれておらず、その科目だけでは単位認定が難しそうである、というようなことも気づきやすくなる。

### 3 ツールを利用して得られた知見

2025年4月に、9つの高専・学科（コース）からの編入学生の単位認定申請のために、シラバス比較用のHTMLファイルの生成をそれぞれ行った。生成した9つのHTMLファイルのサイズは、それぞれにJavaScriptやCSSを含んでいるとはいえ、7.2MB～15.1MBとかなり大きくなってしまった。

#### 3.1 科目間の類似度推定に関して

科目間類似度のスコアは、高いものでは200以上になるものも多くあった。200以上のスコアになる科目のほとんどは数学か物理学に関連する科目で、最も高いスコア(325.2)が算出された科目も微分積分に関する科目であった。

一方、情報科学に関する科目で100以上のスコアになるものはあまり見られなかった。類似度スコアの算出方法より当然のことではあるが、シラバスの記述量がスコアに大きく影響しているようで、図2のように専門用語の種類や出現回数が多い大学の科目は、情報科学に関する科目であっても、類似度スコアが200近くになる科目が各高専にいくつか見つかっている。

シラバスに専門用語の記述が少ない大学の科目は、内容がかなり一致すると思われる高専科目であってもスコアが30程度にしかならないという事例が見られた。とはいえ、他の高専科目とのスコアも低くなるため、内容が一致する科目は上位に表示されることになる。50未満のスコアの科目も表示するようにさえしていれば、類似科目を見逃すことはほぼ無い。

なお、図3は、あまりうまく類似度推定ができていない例である。大学の科目（数値解析）の専門用語の抽出時に用語の区切りが適切に行えていない箇所が多く、もし適切に区切ることができていれば、高専の科目（数値計算）の「ニュートン法」「ガウス」「シン普森」等の専門用語が一致し、前半や中盤のスコアがもう少し高くなっていたはずである。

#### 3.2 ユーザインタフェースに関して

著者の他に、2名の教員がツールを利用して編入生らと単位認定申請の準備を進めていた。申請の準備期間の最中にも、彼らからの意見を取り入れ、以下のようなユーザインタフェースの改良を行った。（図3はその最終形である）

- 大学の科目を授業コード順に並び替えた
- 各項目の開閉状態をブラウザ上 (LocalStorage) に保持するようにした
- 大学と高専のどちらの科目か見分けにくいため、色で区別した（大学→青，高専→緑）
- 用語表示を一つ一つ開くのが手間であったため、デフォルト状態は開いているようにした
- シラバスボタンを押すたびに新しいタブを開くとタブが増えすぎるため改善した
- 大学の科目が必修科目かどうかを表示するようにした

### 4 おわりに

現在のシラバス比較ツールは、高専の卒業年度と学校名・学科・コースを指定して比較するユーザインタフェースとなっている。したがって、編入学生が高専で単位を修得していない選択科目まで表示されてしまう。そこで、編入学生の単位修得した科目と修得した年度を登録する機能、あるいは、未修得科目を非表示にする機能や科目ごとにシラバスデータの年度を変更する機能を追加してみたい。これらの機能を実装できれば、個々の編入学生に合わせたシラバス比較のユーザインタフェースを提供できる。

### 謝辞

本研究の一部においてJSPS 科研費 23K02660の助成を受けた。

### 参考文献

- [1] 国立高等専門学校機構, “高専 Web シラバス”, <https://syllabus.kosen-k.go.jp/> (2025年9月24日閲覧)
- [2] 中川裕志, 前田朗, 小島浩之, “専門用語(キーワード)自動抽出 Python モジュール termextract”, <http://gensen.dl.itc.u-tokyo.ac.jp/pytermextract/> (2025年9月24日閲覧)
- [3] 中川裕志, 湯本紘彰, 森辰則, “出現頻度と連接頻度に基づく専門用語抽出”, 自然言語処理, Vol.10, No.1, pp.27-45, 2003. <https://doi.org/10.5715/jnlp.10.27>