

## 進捗報告，タイトル未定

清水 豪士

**Department of Information Systems Engineering,  
Graduate School of Engineering  
u155016@st.pu-toyama.ac.jp**

**12:10-12:35 Friday, December 10, 2021,  
Toyama Prefectural University.**

- 滝沢の研究ではシラバスからデータを取得する際に、URL を 1 つ 1 つ 入れ替えて、授業計画を取得していた。
  - シラバスのエクセルファイルを取得できた場合のことを考えて、エクセルファイルから授業計画のデータを取得できるようにした。
  - 英語などの教える先生はバラバラだが、授業内容が重複しているものは 1 つにした。
  - 看護学部だけのものや、大学院の授業は除外した。
- 
- 授業計画は担当する先生ごとに書き方が違うため、とりあえず滝沢のを再現するために① ～ ⑮を含む授業計画のものを取得できた。
  - 今後は① ～ ⑮以外の第 1 回や 1 回などといったバラバラのものを取得して作成する

## 取り組んでいること

- 授業計画を基に教材を作成
- Selenium 版と BeautifulSoup4 版の作成（もともと、Selenium で作成）

## Selenium

- ブラウザを自動的に操作するライブラリ.
- 主に WEB アプリケーションのテストや WEB スクレイピング等の自動化するときに用いられる.

## BeautifulSoup

- html ファイルの解析を行うためのライブラリ.
- 解析済みのページのみ処理を行うことができる.
- html データを整形したり, 要素を取得したりする処理を高速に行うことができる.

## Selenium で発生した問題点

- Selenium で教材を作成しているときに、エラーや途中で中断することを繰り返すと reCAPTCHA (bot による Web サイトへの攻撃を防ぐため Google が提供しているサービス) 画面に飛んでしまい、一定期間スクレイピングができない状態になってしまう。



## 対策

- reCAPTCHA 状態になってしまうのを防ぐために、BeautifulSoup でもスクレイピングできるようにする。

- 教材をユーザに薦める際に、youtube の動画や、amazon や楽天などでの本の購入ページを薦めるのは教材として適していないと考える。
- Qiita や Wikipedia などのしっかりと記事が掲載されているものを教材として、提供したい。

- EC サイト系で共通な部分はないか調べる。
- 検索ワードによっては、通販サイトばっかの検索結果になってしまうため、それに対しての対策を考える。

## まとめ

- エクセルファイルから授業計画を作成した.
- Selenium での問題点に対しての対策を考えた.

## 課題

- 授業計画のいろんな書き方に対応する.
- bs4 で教材を作成するためにスクレイピングできるようにする.
- 提供する教材をちゃんと選別する.
- 教務課と話すために、どのようなデータを提供してもらう必要があるのかなどの部分を確認するために、論文をしっかりと理解する.