# IIR Homework 5 方郁文

## 系統環境

* 作業系統 ： Ubuntu 20.04 LTS
* 程式語言 ： Python 3.7.9
* 網站架構 ： Django 3.1.1
* 資料庫： SQLite
* 讀取csv檔 ： Pandas 1.1.3
* 文件來源 : Pubmed Covid-19

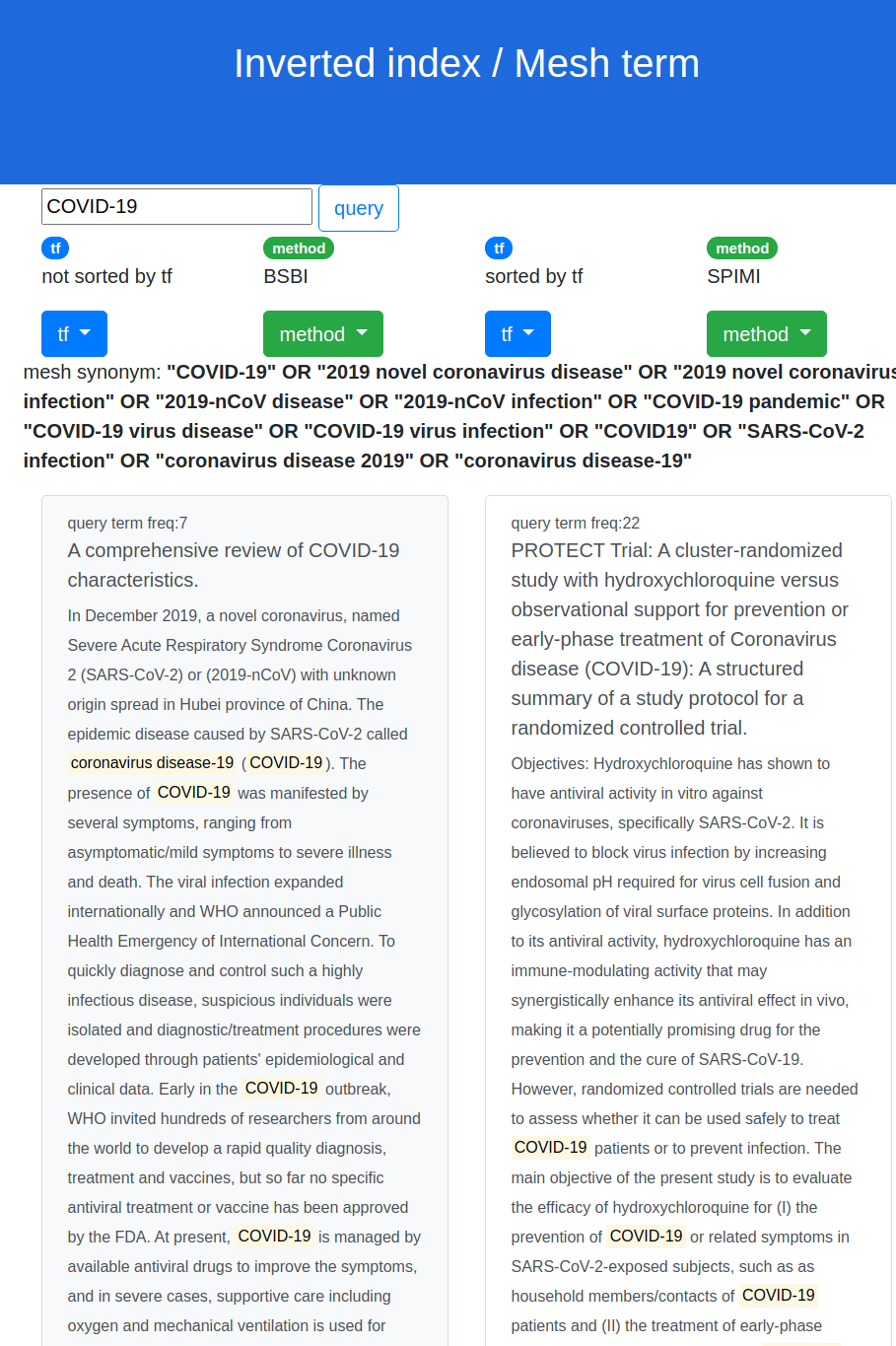
## 實作方法

* 利用python實作BSBI和SPIMI演算法
* 使用作業2、3中使用到的資料集 ( 8434篇Pubmed關於Covid-19的文件 )
* 下載mesh term，將同義詞 ( Synonym ) 建立 dictionary
* 使用mesh term同義詞Dictionary將文件進行tokenize，並分別使用SPIMI和BSBI演算法建立inverted index，比較其費時。
* 搜尋時，將Query Word放入mesh term同義詞Dictionary，得出該字的所有同義詞。
* 將所有同義詞透過inverted index搜尋找到相對應的文件群
* 於前端進行下列幾種比較

1. SPIMI vs BSBI
2. 搜尋出來的文章是否進行根據 tf 排序
3. mesh同義詞搜尋的效果

## 

## 介面展示



tf按鈕: 決定是否依據term frequency進行排列

method按鈕：選擇使用SPIMI或者BSBI 演算法建立的inverted index

## 實驗



上圖為建立分別使用 BSBI 和 SPIMI 建立 inverted index 所需時間