# Web-Based Information Retrieval and Change Detection for Federal Court

**Name:** Canping Liang (canping2)
**Date:** December 11, 2024
**Course:** CS 410 Text Information Systems

## 1. Description

This project might be less directly related to traditional text information retrieval and text mining concepts, but the use of a web crawler has significantly enhanced my work efficiency and provided valuable insights into automating information collection processes. The script developed in this project demonstrates how federal court data can be retrieved, processed, and stored systematically.

The workflow includes **connecting to a database** to manage existing records, requesting data from **online federal court systems**, **processing and comparing retrieved data** with local records to identify discrepancies or updates, and **updating the database** with fresh and validated information. To further streamline the process, a **Python-based scheduler** has been implemented, ensuring that the **script runs automatically at specified intervals**. This automation eliminates the need for manual execution, making the data retrieval process more efficient and reliable.

The source code for this project is included in the folder, accompanied by detailed documentation that outlines each step of the retrieval, update, and scheduling processes. This project highlights the potential of web crawling and scheduling in creating scalable and automated solutions for data-intensive tasks.

## 2. Table of Contents



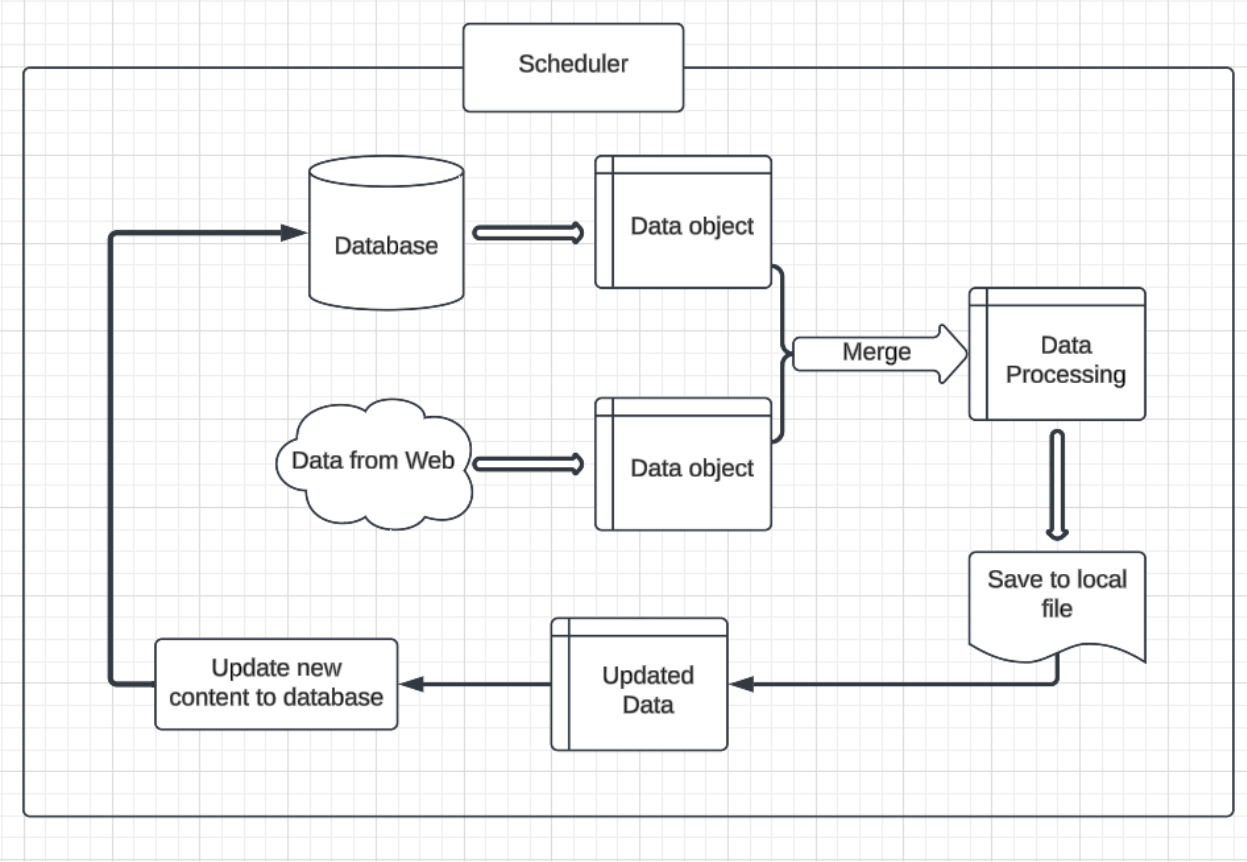
## 3. Tools and Technologies Used

SQL Database Setup
Python 3.10.11 and below libraries are required for running the main class:

```
import psycopg2
import pandas as pd
import requests
import re
import json
import os
from datetime import datetime
```

```
from sqlalchemy import create_engine, text
```

## 4. Basic Layout of the Designed Structure




## 5. Key Steps

1. Connects to the PostgreSQL database.
2. Retrieves data from an online federal court system.
3. Compares the retrieved data with the local database.
4. Updates the database with new or modified records.
5. Saves data logs and outputs in a structured folder.

## 6. Installation

You should be able to download the Python files directly from the provided resources. Please note that this script is specifically designed for a particular use case and requires a custom database setup tailored to the project's requirements. As such, while the script can serve as an example or for informational purposes, it may not run successfully on your own machine without the necessary configurations and dependencies in place.

7. **Presentation**

   **Link to Illinois Media Space**