Grey Literature Scraper: A lightweight Python tool for automated systematic review searching

8 July 2025

Summary

Grey Literature Scraper is a modular Python tool designed to automate the retrieval of grey literature from Ministry of Health websites and Google Scholar. This tool supports systematic reviews in global health, particularly in low- and middle-income countries (LMICs), where critical reports, guidelines, and data are often not available in traditional academic databases.

Statement of Need

Grey literature is essential for comprehensive systematic reviews but is challenging to collect manually due to its dispersed and inconsistent nature. Existing tools often lack modularity or require extensive setup. Grey Literature Scraper addresses these challenges by providing simple, adaptable scripts that enable reproducible, transparent, and efficient searching of government and scholarly grey literature sources.

This tool was developed as part of a systematic review protocol on maternal colonisation with multidrug-resistant Gram-negative bacteria (MDR-GNB) in low- and middle-income countries (LMICs). The review is registered on PROSPERO (CRD420250639803) and the project is openly documented on OSF (https://osf.io/5b78k/?view_only=0c94360369434424b7c19f343513dbdf).

Including this tool in the review process supports transparency, reproducibility, and broad data capture essential for addressing AMR in maternal health in LMIC contexts.

Key Features

- Two modular scrapers:
  - `search_moh2.py` for querying Ministry of Health websites using Google Custom Search API.

– `scholar_scraper3.py` for querying Google Scholar via SerpAPI.
- Customizable queries with user-defined search terms and domains.
- Outputs structured `.csv` files ready for systematic review workflows.
- Focused on LMIC-relevant data sources but extensible for other contexts.

Installation

Clone the repository and install dependencies:

```
git clone https://github.com/lafhlafh/grey-literature-scraper.git
cd grey-literature-scraper
python3 -m venv envs/greyenv
source envs/greyenv/bin/activate   # macOS/Linux
pip install -r requirements.txt
```

Usage

Run either scraper script:

```
python search_moh2.py     # Search Ministry of Health websites
python scholar_scraper3.py  # Search Google Scholar
```

Configure API keys as described in the README.

Authors

Lauren Hookham, PhD Candidate, Brighton and Sussex Medical School & MRC/UVRI & LSHTM Uganda Research Unit E- mail: l.hookham1@uni.bsms.ac.uk

Acknowledgements

License

MIT License — see LICENSE file.