🚀 OneCLick Keyword Retriever

OneCLick is a lightweight, open-source Streamlit application designed to enrich bibliometric datasets by retrieving missing keywords from the OpenAlex API. It supports streamlined preprocessing for tools such as VOSviewer, CiteSpace, and Bibliometrix, and provides reproducible export formats for downstream analysis.

Features

- Upload Input: Excel file (.xlsx) with a column named DI (DOI).
- Request–Response Flow:
  - Sends DOI-based queries to OpenAlex.
  - Retrieves enriched metadata keywords.
- Error & Warning Handling:
  - Invalid DOIs, network timeouts, or rate-limit failures logged in Failed_DOIs.csv.
  - Valid DOIs with no keywords flagged as [WARNING:NoKeywords] and logged in NoKeyword_DOIs.csv.
- Latency Logging:
  - Per-DOI request latency is measured.
  - Median and IQR statistics are computed.
  - Full log available as Latency_Log.csv.
- Outputs:
  - Enriched dataset (openalex_keywords.csv / .xlsx) including an OpenAlex_KW column.
  - Separate downloadable error/warning logs.
  - Latency log for reproducibility and benchmarking.

Installation

Clone the repository and install dependencies:

```
git clone https://github.com/meMeta-a11y/oneclick.git
cd oneclick
pip install -r requirements.txt
```

Run the Streamlit app:

```
streamlit run streamlit_app.py
```

Example Datasets

To support reproducibility and benchmarking, three public test datasets are provided:

- Pilot (37 DOIs) – for small-scale exploratory testing.
- Intermediate (100 DOIs) – for lightweight benchmarking.
- Validation (1,097 DOIs) – for large-scale robustness evaluation.

These are maintained on the project's GitHub repository and permanently archived via Zenodo.

Performance Benchmarking

Validation experiments confirm OneCLick's robustness and scalability.

Scale of Dataset	Number of DOIs	Error DOIs	No-Keyword DOIs	Successful Enrichment (%)	Median Latency (s/DOI, IQR)*
Pilot	37	0	0	100.00%	0.30 seconds (IQR: 0.28 – 0.31 seconds)
Intermediate	100	0	26	74.00%	0.29 seconds (IQR: 0.28 – 0.31 seconds)
Validation	1,097	0	205	81.31%	0.29 seconds (IQR: 0.28 – 0.30 seconds)

*Benchmarks were conducted on a Streamlit Cloud deployment in August 2025, using a Linux machine (4 GB RAM, 100 Mbps internet). Runtime and error rates may vary depending on network conditions and OpenAlex API rate limits. The lowest median latency per DOI (inter-quartile range, IQR) across both repeats is reported.

Import Quick-Start (VOSviewer Example)

1. Open VOSviewer → Create → Create a map based on network data.
2. Select Read from bibliometric file.
3. Choose the OneCLick-enriched .csv.
4. In the import wizard, map OpenAlex_KW as the keyword field.
5. Select options for co-occurrence counting (e.g., full counting).
6. Generate and visualize the keyword network.

Parallel workflows are documented in the manuscript (Section 2.12).

Limitations and Future Work

- Topic hierarchies and confidence scores are not yet exported; these are reserved for future development.
- Integration into bibliometric clustering pipelines (e.g., theme assignment) is under development.

Repository Structure

```
oneclick/
│
├── streamlit_app.py          # Main application
├── requirements.txt          # Dependencies
├── example_data/             # Example DOI datasets
│   ├── pilot_37.xlsx
│   ├── intermediate_100.xlsx
│   └── validation_1097.xlsx
├── outputs/                  # Example enriched outputs + logs
│   ├── openalex_keywords.csv
│   ├── Failed_DOIs.csv
│   ├── NoKeyword_DOIs.csv
│   └── Latency_Log.csv
└── README.md                 # Documentation
```

Citation

If you use OneCLick in your research, please cite:

> *If you use OneCLick in your research, please cite it as: WEI, Loo Keat. (2025). OneCLick: Streamlined Metadata Enrichment Using Machine-Inferred Keywords from OpenAlex (Version 1.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.15876773.*
