# PubMedXport: Export PubMed Metadata and Abstracts to CSV, JSON, and BibTeX

**Myo Kyi Tha** [1] and **Nilar Khin** [1]

**1** MAHSA University, Malaysia ROR

abstract: | Efficient extraction of structured metadata from biomedical literature is critical for systematic reviews, meta-analyses, and bibliometric studies. PubMed, while central to biomedical research, offers limited export capabilities. We present PubMedXport, a web-based tool that enables users to search PubMed and export results, including abstracts, in CSV, JSON, and BibTeX formats. Featuring an intuitive interface, customizable queries, parallel processing, and error reporting, PubMedXport streamlines data retrieval for systematic reviews, meta-analyses, and educational purposes. Developed through a combination of human expertise and AI-assisted coding, it exemplifies the accelerating role of AI in research tool development.

long-description: | **Summary**

Efficient literature searching underpins evidence-based practice. PubMed, operated by the U.S. National Library of Medicine, offers robust search capabilities but limited export functionality: abstracts are excluded from CSVs, structured formats like JSON are unavailable, and BibTeX entries must be manually generated. Abstracts are only available as plain, unstructured text, hindering efficient data use.

PubMedXport addresses these gaps through a lightweight, browser-based application that retrieves full metadata, including abstracts, and exports results in structured formats without requiring installation. Supporting custom queries, year filtering, parallel processing, and error reporting, PubMedXport facilitates systematic reviews, meta-analyses, and educational research.

**Background**

PubMed provides free, high-quality access to MEDLINE and other biomedical resources, distinguishing itself from subscription-based platforms like Scopus and Web of Science. Despite its powerful search capabilities, its data export options are constrained, especially for structured or large-scale computational workflows. Tools that extend PubMed's export features are needed to meet the growing demands of systematic reviews and computational analysis.

Statement of Need: |

Researchers face several barriers when using PubMed for structured data collection:

- Abstracts are excluded from CSV exports.
- JSON format is unavailable.
- BibTeX citations require manual generation.
- Abstracts are supplied only as unstructured text files.

While third-party tools exist, they often require local installation or complex workflows. PubMedXport overcomes these limitations by offering a browser-based, user-friendly platform for exporting PubMed search results, including abstracts, in multiple formats.

**Features**

- **PubMed Search**: Custom keyword queries with optional year range filters.
- **Data Retrieval**: Full metadata and abstract collection.

- **Author Country Extraction**: Extract country-level info for all authors (First author, and co-authors, if any, in separate fields - with duplications removed).
- **Export Options**: CSV (with abstracts), JSON (machine-readable), BibTeX (citation-ready).
- **Parallel Requests**: Faster data retrieval through simultaneous API calls.
- **Progress Tracking**: Real-time indicators during searches.
- **Error Handling**: Automated listing of failed PMIDs.
- **Browser-Based**: No installation required.

**Usage**

1. Visit PubMedXport.
2. Enter your PubMed API key and define query parameters.
3. Start the search and download results.
4. Review any retrieval errors from the provided .txt file.

**How to get the code**

Clone or download from GitHub.

**Example screenshots of the user interface:**

- Result with no failed PMID: Link
- Result with 1 failed PMID: Link
- Result on iOS: Link
- CSV Output: Link
- JSON Output: Link
- BibTeX Output: Link

**Development**

PubMedXport was created using a hybrid approach that blended human expertise with free AI tools (ChatGPT, GitHub Copilot, and DeepSeek) for coding and troubleshooting. The authors designed the concept, chose the features, performed extensive testing, and focused on user-centered design, demonstrating how AI-augmented development can enhance research tool creation.

**Limitations and Future Directions**

While PubMedXport extends PubMed's capabilities, it remains dependent on PubMed's data structures and API limits.

Planned enhancements include:

- **Automated Abstract Summarization**: Integration of BioBART, Facebook BART, or other models for rapid abstract review.
- **Research Relevance Check**: NLP-based assessment of article relevance to user-defined queries.
- **Expanded Export Options**: Additional formats (e.g., XML, TSV) and advanced filtering features.

**Author Contributions**

Myo Kyi Tha and Nilar Khin contributed equally to this work.

**Acknowledgments**

**License**

Distributed under the MIT License.

**References**

1. National Library of Medicine. PubMed Overview. Available at: https://pubmed.ncbi.nlm.nih.gov/about/
2. National Center for Biotechnology Information (NCBI). E-utilities API Documentation. Available at: https://www.ncbi.nlm.nih.gov/books/NBK25500/