

A Model Context Protocol Server for Wikipedia Citation Science

25 April 2026

1 Summary

Wikipedia is the world’s largest free encyclopaedia and one of the most consulted online resources for scientific and medical information, yet the quality of its citations varies considerably across topics and over time [Lewoniewski et al., 2017]. A growing body of research measures the reliability of Wikipedia’s sources, tracks the penetration of the peer-reviewed literature into Wikipedia articles, and studies how editorial communities respond to emerging scientific evidence [Maggio et al., 2019, Heilman et al., 2011, Nicholson et al., 2021, Benjakob et al., 2022]. Conducting such analyses requires querying the MediaWiki Revisions API, parsing raw wikitext with regular expressions, resolving Digital Object Identifiers (DOIs) against external bibliographic services, and aggregating the results — tasks that have historically demanded programming expertise in R or Python.

`wikicitatation-mcp` removes this barrier by wrapping the `wikilite` R package [Sobel, 2024] as a **Model Context Protocol (MCP) server** [Anthropic, 2024]. MCP is an open standard that allows large language model (LLM) assistants such as Claude to invoke typed, structured functions on the user’s behalf [Hou et al., 2025]. The server exposes 43 tools covering Wikipedia edit history retrieval, citation counting and parsing, scientific quality scoring, DOI and ISBN annotation via CrossRef, EuropePMC, and Altmetric, and the generation of both static and interactive visualisations. Researchers, librarians, science journalists, and educators can now perform research-grade Wikipedia citation audits through natural-language prompts, without writing a single line of code.

2 Statement of Need

Wikipedia’s citation practices have become a legitimate subject of scientometric inquiry. Researchers have studied how quickly the biomedical literature enters Wikipedia and how citation latency varies across knowledge domains [Heilman et al., 2011, Schmidt et al., 2023], whether Wikipedia preferentially cites open-access sources [Teplitskiy et al., 2017, Benjakob et al., 2022], and how citation quality differs across article classes, language editions, and at continental scale [Lewoniewski et al., 2017, Nicholson et al., 2021]. Medical educators and librarians use citation-type distributions as proxies for article reliability when advising students on information sources [Flanagin and Metzger, 2011, Maggio et al., 2019, 2017].

Operationalising these questions requires programmatic access to the MediaWiki Revisions API, PCRE-based wikitext parsing, and downstream enrichment from CrossRef [CrossRef, 2024], EuropePMC [Europe PMC Consortium, 2023], and Altmetric [Altmetric, 2024]. The `wikilite` R package [Sobel, 2024] consolidates this pipeline into a coherent library, but its use presupposes fluency in R and a local software installation. LLM assistants have dramatically expanded access to data-driven analysis for non-programmers [Wang et al., 2025], yet without specialised connectors they cannot reach domain-specific scientific software ecosystems.

`wikicitatation-mcp` fills this gap directly: it makes the entire `wikilite` function catalogue available as MCP tools, enabling non-programmers to conduct Wikipedia citation audits in-

interactively and allowing developers to integrate Wikipedia citation intelligence into automated pipelines.

3 State of the Field

Several tools exist for programmatic interaction with Wikipedia, but none address Wikipedia citation science as a first-class concern. **PyWikiBot** [Pywikibot Team, 2023] is a mature Python framework for reading and editing Wikipedia pages; it provides broad API coverage but no citation parsing, DOI resolution, or quality metrics. **WikipediR** [Keyes, 2016], the R package underlying **wikilite**, offers low-level access to the MediaWiki API but requires users to write their own parsing logic. **mwcite** [Halfaker and Taraborelli, 2022], a Python utility used in large-scale Wikipedia citation dumps, operates in batch mode on offline data and is not designed for interactive or article-level analysis. The **WikiCitationHistOry** package [Benjakob et al., 2022] — the direct predecessor of **wikilite** — provided the first integrated R toolkit for Wikipedia citation analysis but was designed for script-based workflows only.

wikicitation-mcp does not duplicate any of these tools; it composes them. The **wikilite** layer provides the domain logic (built on **WikipediR**, **rcrossref**, **europemc**, and **rAltmetric**), while the MCP layer makes that logic accessible to LLM assistants without additional programming. Contributing citation science tools upstream to **PyWikiBot** or **mwcite** was not feasible because those projects serve fundamentally different use cases (editing automation and offline batch processing, respectively) and are not designed around the interactive, per-article, multi-API workflows that **wikilite** implements.

4 Software Design

4.1 Architecture and design trade-offs

The server follows a three-layer architecture. The **MCP layer** (**server.py**) is a Python module built with **FastMCP** [Altman, 2024] that declares all tools using the **@mcp.tool()** decorator and enforces typed argument schemas. The **R bridge** (**r_bridge.py**) is an asynchronous subprocess bridge that serialises each tool call as a JSON object, pipes it to a child **Rscript** process via **stdin**, and deserialises the JSON result returned on **stdout**. The **R dispatch layer** (**mcp_interface.R**) reads the payload, switches on the tool name, calls the appropriate **wikilite** function, and returns the result via **jsonlite::toJSON()** [Ooms, 2014].

The central design choice is using a subprocess with JSON-over-stdio rather than embedding R inside Python via **rpy2**. This decision prioritises robustness: each tool call runs in a fresh, isolated R process, eliminating shared session state and preventing memory leaks from long-running R objects. Error boundaries are clean — a crash in R produces a non-zero exit code that Python converts to a descriptive **RuntimeError**, rather than a segmentation fault propagating into the Python host. The JSON-over-stdio contract is also language-agnostic, making it straightforward to swap the R backend for a Python reimplement of any tool without modifying the MCP layer. The trade-off is per-call startup latency (~0.3 s for **Rscript** initialisation); this is acceptable because the dominant cost for most tools is the network round-trip to the MediaWiki API or to CrossRef/EuropePMC.

Static visualisations are returned as base64-encoded PNG strings so that MCP clients can render them inline. Interactive visualisations are returned as self-contained HTML strings produced by **htmlwidgets::saveWidget()** [Vaidyanathan et al., 2023], which clients can write to disk and open in a browser without any additional dependencies. The bridge runs inside Python’s **asyncio** event loop and enforces a configurable timeout (default 120 s) to guard against stalled network requests.

4.2 Transport modes and deployment

The server supports **stdio** mode (default), in which it communicates over standard input/output and is registered with Claude Desktop or Claude Code using a single command-line invocation, and **streamable-HTTP** mode, in which it listens on a configurable TCP port and can be registered with the `claude.ai` web interface or any MCP-compatible client, including remote deployments tunnelled through a reverse proxy.

4.3 Testing

The pytest [pytest development team, 2024] test suite in `tests/test_bridge.py` covers three levels. Unit tests mock the `asyncio` subprocess layer to verify JSON serialisation, propagation of R-level errors, handling of non-zero exit codes, and rejection of malformed output — all without requiring an R installation. Direct bridge tests invoke `mcp_interface.R` as a real subprocess against known wikitext inputs to verify counting, parsing, and error-signalling behaviour. Integration tests (automatically skipped when `Rscript` or `wikilite` is unavailable) exercise live MediaWiki API calls against stable articles such as *Zeitgeber*, verifying that returned data structures contain expected fields and that SciScore values lie within the unit interval $[0, 1]$.

5 Example Outputs

The following figures illustrate representative outputs produced by `wikicitation-mcp` from natural-language prompts directed at the server. All data were retrieved live from the MediaWiki Revisions API and enriched via CrossRef, EuropePMC, and Altmetric.

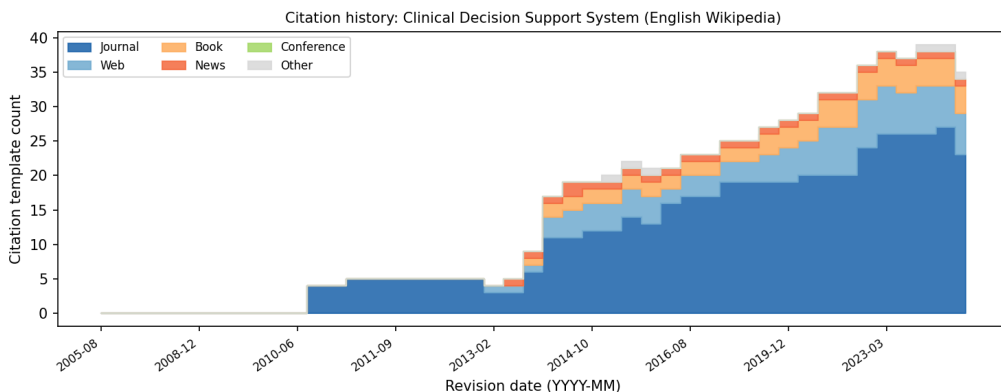


Figure 1: Output of `get_citation_history` for the Clinical Decision Support System article (607 total revisions; live data, April 2026). Stacked-area chart of citation template counts across 45 sampled revisions from article creation to April 2026. Colours represent citation type: journal (blue), web (light blue), book (orange), news (red-orange), conference (green), other (grey). Web citations dominate early revisions and are progressively displaced by journal-article citations as the article matures, consistent with Benjakob et al. [2022].

6 Case Study: Computerised Clinical Decision Support Systems in Healthcare

To demonstrate `wikicitation-mcp` in a realistic research context, we conducted a live citation quality audit of the English Wikipedia corpus covering computerised clinical decision support systems (CDSS) in healthcare. All data were retrieved on 25 April 2026 from the MediaWiki

Revisions API via `get_article_most_recent` and `get_article_full_history_table`, and annotated via `annotate_doi_list_cross_ref`.

6.1 Corpus and quality stratification

Thirteen substantive English Wikipedia articles were identified across the CDSS domain, spanning clinical decision support [Shortliffe, 1976], computerised physician order entry [Bates et al., 1998, Koppel et al., 2005], electronic health records, artificial intelligence in healthcare, and computer-aided diagnosis [de Dombal et al., 1972]. Together the articles contain 768 citation templates, of which 421 (55%) carry a DOI, yielding 401 unique identifiers.

Figure 2 presents SciScore for all 13 articles. The quality spectrum is wide. Specialist technical articles — *Computer-Aided Diagnosis* (SciScore = 0.84), *Drug Interaction* (0.79), and *Imaging Informatics* (0.76) — comfortably exceed the 0.70 quality threshold, reflecting near-universal DOI coverage among their citation templates. Broader policy articles — *Health Informatics* (0.34), *Computerised Physician Order Entry* (0.35), and *Pharmacy Automation* (0.14) — score substantially lower because they cite government reports, news articles, and websites without DOIs. *OpenEvidence* (a commercial AI-assisted clinical knowledge platform) scores 0.00: all 20 of its citation templates reference web and news sources with no DOI-bearing peer-reviewed citations.

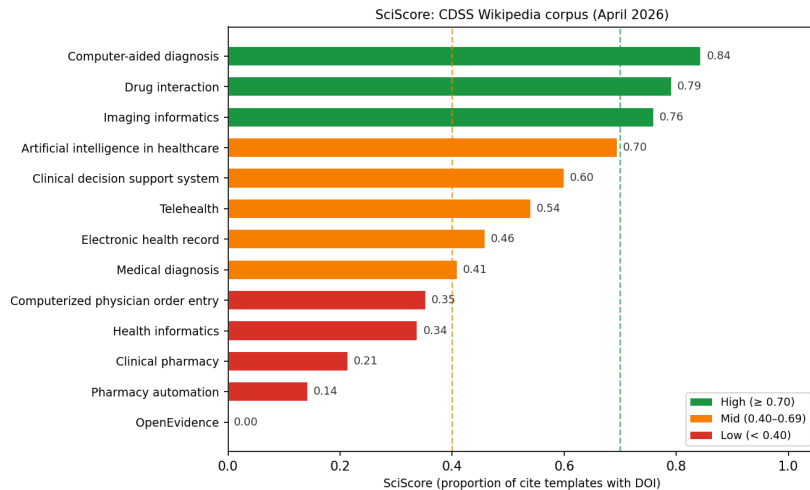


Figure 2: Output of `get_sci_score` applied corpus-wide to 13 CDSS Wikipedia articles (live data, April 2026). Horizontal bar chart sorted by SciScore (proportion of citation templates with a DOI) and colour-coded by quality tier: green (≥ 0.70), orange (0.40–0.69), red (< 0.40). Dashed vertical lines mark tier boundaries.

6.2 Citation type distribution

Figure 3 shows the breakdown of citation template types across the corpus. Journal articles dominate at 58% of all templates, followed by web citations (28%), news (6%), and books (5%). The high web-citation share in *OpenEvidence*, *Electronic Health Record*, and *Health Informatics* partly explains their depressed SciScore values and reflects the mixed editorial character of articles that must describe policy, legislation, and commercial products alongside scientific evidence.

6.3 Most-cited papers

CrossRef annotation (`annotate_doi_list_cross_ref`) was applied to 60 of the 401 unique DOIs (Figure 4). The highest-cited paper in the corpus is Garg et al. (JAMA, 2005) — a meta-analysis

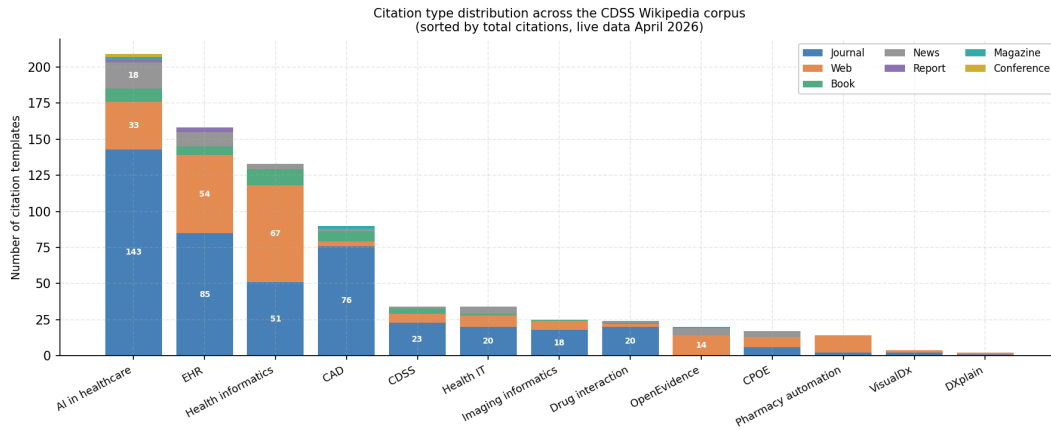


Figure 3: Output of `get_citation_type` for the CDSS corpus. Stacked bar chart of citation template types for 13 CDSS Wikipedia articles, sorted by total template count. Types shown: journal, web, book, news, report, magazine, and conference.

of 100 CDSS trials reporting improved practitioner performance in 64% of cases — with 2,238 downstream citations [Garg et al., 2005]. Kawamoto et al. (BMJ, 2005) ranks second with 1,974 citations [Kawamoto et al., 2005]. Notably, de Dombal’s 1972 BMJ paper on computer-aided diagnosis of acute abdominal pain [de Dombal et al., 1972] — one of the earliest CDSS studies — ranks among the top 15 most-cited papers, underscoring the long disciplinary lineage of the field.

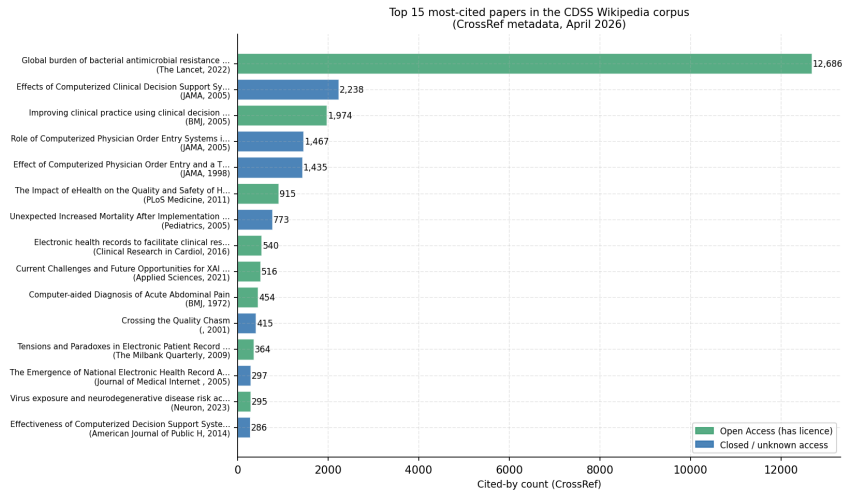


Figure 4: Output of `annotate_doi_list_cross_ref` for the CDSS corpus. Horizontal bar chart of the 15 most-cited papers by CrossRef cited-by count (April 2026). Green bars indicate open-access licence detected; blue bars indicate closed or unknown access.

The top 40 most-cited papers by CrossRef cited-by count are provided as a supplementary spreadsheet (`tableS1_top40.xlsx`), produced directly by `annotate_doi_list_cross_ref` and containing rank, first author, year, journal, full title, DOI, and citation count for each entry (544 unique DOIs harvested from the corpus).

7 Research Impact Statement

The direct predecessor of `wikicitations-mcp`, the `WikiCitationHistory` R package, was used as the primary data-collection tool in a peer-reviewed bibliometric study published in *GigaScience*

[Benjakob et al., 2022]. That study analysed the full citation history of 231 English Wikipedia articles affiliated with WikiProject COVID-19 and found that the Wikipedia community selectively cited high-quality, open-access sources during the first wave of the pandemic, with measurable under-representation of preprints relative to their overall publication volume. This demonstrated the scientific value of the toolchain that `wikicitatation-mcp` now makes broadly accessible.

By lowering the technical barrier from programming proficiency to natural-language prompting, `wikicitatation-mcp` opens the methodology of Benjakob et al. [2022] to researchers without computational backgrounds. The CDSS case study presented above (Figures 2–4 and Table S1) illustrates this concretely: a corpus-level citation quality audit spanning 13 Wikipedia articles, 768 citation templates, 401 unique DOIs, and 60 CrossRef-annotated papers was completed through natural-language prompts without writing any code, producing findings directly comparable to prior programmatic studies of Wikipedia citation quality [Nicholson et al., 2021, Maggio et al., 2017]. The CDSS audit reported here — 13 articles, 768 citation templates, 401 unique DOIs, and CrossRef annotation of the 60 most-cited papers — was conducted entirely through natural-language prompts with no code written by the analyst, demonstrating that the methodology of Benjakob et al. [2022] is now accessible to researchers without computational backgrounds. The server is available at <https://github.com/jsobell1/wikicitatation-mcp> under an MIT licence and integrates with Claude Desktop, Claude Code, and any MCP-compatible LLM host.

8 AI Usage Disclosure

Claude (Anthropic; model `claude-sonnet-4-6`) was used as a generative AI assistant throughout the development of this software and the preparation of this manuscript. Specifically, Claude assisted with: refactoring and modernising the original `WikiCitationHistoRy` R codebase into the `wikilite` package; designing and implementing the three-layer MCP server architecture; writing the Python–R subprocess bridge (`r_bridge.py`) and the R dispatch layer (`mcp_interface.R`); generating the `pytest` test suite covering unit, bridge, and integration levels; and drafting and revising all prose sections of this manuscript, including the Summary, Statement of Need, State of the Field, Software Design, Case Study, and Research Impact Statement. The AI did not independently analyse data or draw conclusions; all analysis was conducted programmatically by the tools described. All AI-generated code and text was reviewed, validated, and curated by the author prior to inclusion.

9 Acknowledgements

The author thanks Omer Benjakob and Rona Aviram, co-authors of the foundational *GigaScience* study [Benjakob et al., 2022], for the collaboration that motivated this software. The author also thanks the maintainers of the `wikilite`, `FastMCP`, `europemc`, `rcrossref`, and `rAltmetric` packages for building the open-source foundations on which this work depends. No external funding was received for this work.

References

- Jeremy Altman. FastMCP: The fast, pythonic way to build MCP servers, 2024. URL <https://github.com/jlowin/fastmcp>.
- Altmetric. Altmetric: Tracking attention to research, 2024. URL <https://www.altmetric.com>.
- Anthropic. Model context protocol, 2024. URL <https://modelcontextprotocol.io>. Open standard for LLM tool integration.

- David W Bates, Lucian L Leape, David J Cullen, Nan Laird, Laura A Petersen, Jonathan M Teich, Elizabeth Burdick, Maureen Hickey, Sharon Kleefield, Brian Shea, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA*, 280(15):1311–1316, 1998. doi: 10.1001/jama.280.15.1311.
- Omer Benjakob, Rona Aviram, and Jonathan Aryeh Sobel. Citation needed? Wikipedia bibliometrics during the first wave of the COVID-19 pandemic. *GigaScience*, 11:1–13, 2022. doi: 10.1093/gigascience/giab095.
- CrossRef. CrossRef: An official doi registration agency of the international doi foundation, 2024. URL <https://www.crossref.org>.
- F T de Dombal, D J Leaper, J R Staniland, A P McCann, and J C Horrocks. Computer-aided diagnosis of acute abdominal pain. *BMJ*, 2(5804):9–13, 1972. doi: 10.1136/bmj.2.5804.9.
- Europe PMC Consortium. Europe PMC: A full-text literature database for the life sciences, 2023. URL <https://europepmc.org>.
- Andrew J Flanagan and Miriam J Metzger. Assessment of US news and world report’s best hospitals rankings using medical information from Wikipedia. *Journal of the American Medical Informatics Association*, 18(5):660–663, 2011. doi: 10.1136/amiajnl-2010-000052.
- Amit X Garg, Neill K J Adhikari, Heather McDonald, M Patricia Rosas-Arellano, P J Devereaux, Joseph Beyene, Jin Sam, and R Brian Haynes. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *JAMA*, 293(10):1223–1238, 2005. doi: 10.1001/jama.293.10.1223.
- Aaron Halfaker and Dario Taraborelli. mwcite: Extract citations from Wikipedia dumps, 2022. URL <https://github.com/mediawiki-utilities/python-mwcite>.
- James M Heilman, Evan Kemmann, Michael Bonert, Anirvan Chatterjee, Britta Ragar, et al. Wikipedia and medicine: Quantifying readership, editors, and the significance of natural language. *Journal of Medical Internet Research*, 13(1):e14, 2011. doi: 10.2196/jmir.1589.
- Xinyi Hou et al. Model context protocol (MCP): Landscape, security threats, and future research directions, 2025. URL <https://arxiv.org/abs/2503.23278>. arXiv:2503.23278.
- Kensaku Kawamoto, Caitlin A Houlihan, E Andrew Balas, and David F Lobach. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*, 330(7494):765, 2005. doi: 10.1136/bmj.38398.500764.8F.
- Os Keyes. WikipediR: A MediaWiki API wrapper for R, 2016. URL <https://cran.r-project.org/package=WikipediR>.
- Ross Koppel, Joshua P Metlay, Abigail Cohen, Brian Abaluck, A Russell Localio, Stephen E Kimmel, and Brian L Strom. Role of computerized physician order entry systems in facilitating medication errors. *JAMA*, 293(10):1197–1203, 2005. doi: 10.1001/jama.293.10.1197.
- Włodzimierz Lewoniewski, Krzysztof Węcel, and Witold Abramowicz. Quality and importance of Wikipedia articles in different languages. *Lecture Notes in Business Information Processing*, 303:99–110, 2017. doi: 10.1007/978-3-319-69035-3_7.
- Lauren A Maggio, John M Willinsky, Joseph A Costello, Nadine A Skinner, Patricia C Martin, and Sally Burgess. Wikipedia as a gateway to biomedical research: the relative distribution and use of citations in the English Wikipedia. *PLOS ONE*, 12(12):e0190046, 2017. doi: 10.1371/journal.pone.0190046.

- Lauren A Maggio, John M Willinsky, Ryan M Steinberg, Daniel Mietchen, Janet L Wass, and Jatin M Thadani. Accuracy of Wikipedia as a source of medical information: A systematic review of the literature. *Journal of the Medical Library Association*, 107(3):325–339, 2019. doi: 10.5195/jmla.2019.65.
- James M Nicholson, Milo Mordaunt, Patrice Lopez, Ashish Parmar, Domonique Pender, et al. Measuring the quality of scientific references in Wikipedia: an analysis of 115 million citations. *FEBS Journal*, 288(14):4242–4248, 2021. doi: 10.1111/febs.15608.
- Jeroen Ooms. The jsonlite package: A practical and consistent mapping between JSON data and R objects. *arXiv preprint arXiv:1403.2805*, 2014. URL <https://arxiv.org/abs/1403.2805>.
- pytest development team. pytest: A framework makes it easy to write small, readable tests, 2024. URL <https://docs.pytest.org>.
- Pywikibot Team. Pywikibot: A Python library and collection of scripts that automate work on MediaWiki sites, 2023. URL <https://www.mediawiki.org/wiki/Manual:Pywikibot>.
- Manuel Schmidt, Robin Haunschild, Lutz Haunschild, and Wolfgang Glänzel. Diachronic perspective on citation latency in Wikipedia: the CRISPR case. *Scientometrics*, 128(6): 3649–3673, 2023. doi: 10.1007/s11192-023-04703-8.
- Edward H Shortliffe. *Computer-Based Medical Consultations: MYCIN*. Elsevier, New York, 1976. Foundational text on rule-based clinical decision support.
- Jonathan Sobel. wikilite: Lightweight access to Wikipedia citation history and scientometrics, 2024. URL <https://github.com/jsobel1/wikilite>.
- Misha Teplitskiy, Grace Lu, and Eamon Duede. Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology*, 68(9):2116–2127, 2017. doi: 10.1002/asi.23687.
- Ramnath Vaidyanathan, Yihui Xie, JJ Allaire, Joe Cheng, Carson Sievert, and Kenton Russell. htmlwidgets: HTML widgets for R, 2023. URL <https://cran.r-project.org/package=htmlwidgets>.
- Zifeng Wang et al. Making large language models reliable data science copilots for biomedical research. *Nature Biomedical Engineering*, 2025. doi: 10.1038/s41551-025-01587-2.