

Workshop on Citation Extraction and Parsing (CiteX 2026)

28. and 29. May 2026 | Frankfurt a. M. (Germany)

Scope

Accurate and open citation data are essential foundations for transparent, reproducible, and networked scholarship. The [Workshop on Citation Extraction and Parsing \(CiteX 2026\)](#) provides an interdisciplinary forum for researchers, developers, and practitioners to explore current advances in the automated identification, structuring, and dissemination of bibliographic references.

Building upon the growing momentum around open citations and initiatives such as [WikiCite](#), [WOOC](#), and the Frankfurt Workshop series “[New Approaches for Extracting Heterogeneous Reference Data](#)”, CiteX 2026 aims to foster dialogue across communities working on both the methodological and infrastructural aspects of citation data. The workshop welcomes contributions addressing technical innovations—ranging from traditional rule-based and machine-learning methods to state-of-the-art approaches using large language models (LLMs)—as well as practical and conceptual discussions about the creation, quality, and integration of open citation datasets.

CiteX 2026 invites participation from all disciplines interested in the extraction, parsing, and reuse of citation information. By bringing together perspectives from research, infrastructure, and applied contexts, the workshop seeks to strengthen collaboration, identify shared challenges, and promote the development of interoperable and openly accessible citation ecosystems.

Topics of Interest include, but are not limited to:

- Automated extraction and parsing of references
- Creation and sharing of gold standards and test datasets
- Standardization and interoperability of citation data
- Quality assessment and validation of extracted references
- Provision and integration of open citation data into repositories and search systems
- Citation practices across disciplines
- Data linking between scholarly works, datasets, and other research outputs
- Annotation and enrichment of citation data
- Prompt engineering and fine-tuning of LLMs (e.g., GPT-4, LLaMA) for citation tasks
- Comparison of LLM-based and tool-based (e.g., GROBID, Anystyle, Cermine) extraction pipelines
- In-text citation extraction and context analysis using LLMs
- Use of open web search APIs or LLMs for source retrieval

Contributions and registration

We are looking forward to receiving submissions for presentations, poster and hands-on sessions.

We do prefer presenters to participate on-site; however we try to make online presentations possible. Please indicate if you will attend on-site or online in the submissions form.

Participation is as well possible without contribution. We will charge a small fee to cover the costs for organisation. More information on the registration will follow in mid-November.

Submission guidelines

- Please submit an extended abstract (1250-1500 words excl. references) for any contribution format.
- Only submissions in English will be considered.
- Accepted extended abstract will be published under a Zenodo community.
- Please indicate in the submission form, if you are interested in a Special Issue.

Important dates

- Submission deadline: 15. January 2026
- Notification of acceptance: 01. March 2026
- Camera-ready version: 31. March 2026
- Workshop dates: 28./ 29. May 2026

Venue

- The Workshop will take place at DIPF | Leibniz Institute of Research and Information in Education, Rostocker Straße 6, 60323 Frankfurt a. M.

Submission & Contact

- Please submit your contribution via Google Forms on our website. More information will follow in mid-November.
- The workshop website is available under: [Workshop on Citation Extraction and Parsing](#)
- For any question, please contact us via e-mail: workshop.cep.2026@gmail.com

Programme Committee

- | | |
|--|--|
| • Tamara Heck (DIPF Leibniz Institute for Research and Information in Education) | • Marta Soricetti (University of Bologna) |
| • Angelo Di Iorio (University of Bologna) | • Matteo Romanello (Odoma) |
| • Philipp Mayr-Schlegel (GESIS – Leibniz Institute for the Social Sciences) | • Christian Boulanger (Max Planck Institute for Legal History and Legal Theory) |
| • Silvio Peroni (OpenCitations) | • Andreas Wagner (Max Planck Institute for Legal History and Legal Theory) |
| • Christoph Schindler (DIPF Leibniz Institute for Research and Information in Education) | • Daniel Mietchen (Freie Universität Berlin) |
| • Stephan Stahlschmidt (German Centre for Higher Education Research and Science Studies) | • Muhammad Ahsan Shahid (GESIS – Leibniz Institute for the Social Sciences) |
| | • Verena Weimer (DIPF Leibniz Institute for Research and Information in Education) |

The workshop is part of and organised by the project team of OFFZIB - Open Citation Data for Educational Research, a project funded by the *German Research Foundation*.