



WORKSHOP PROGRAMME

Reference Extraction at the Intersection of AI Research and the Digital Humanities: Validation, Interoperability and Collaboration

04.11.2025 | 09:00-18:30 | MPILHLT (HYBRID)

Convenors: Christian Boulanger (mpilhlt), Andreas Wagner (mpilhlt)

Tuesday, 04.11.2025

I. ONBOARDING

09:00–09:15 **Arrival/Registration**

09:15–09:20 **Welcome**

Christian Boulanger, Andreas Wagner (mpilhlt)

09:20–09:45 **Upshot from RefExtract2023, State of the Discussion**

Christian Boulanger, Andreas Wagner (mpilhlt)

09:45–10:00  Coffee

II. RESEARCH

10:00–12:30 **In-depth Research Impact Summarization through Fine-Grained Temporal Citation Analysis**

Hiba Arnaout (TU Darmstadt)

Benchmarking Large Language Models on Reference Extraction and Parsing in the Social Sciences and Humanities

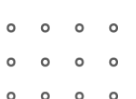
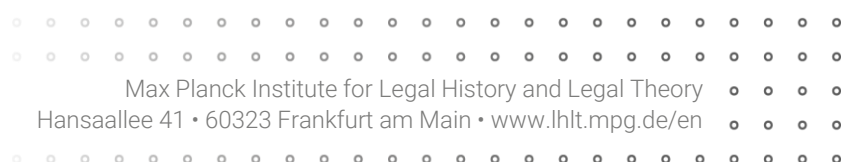
Yurui Zhu (Odoma), Matteo Romanello (Odoma)

How Scientific Ideas Evolve

Sofía Aguilar Valdez (Saarland University)

Open Discussion and Ad-Hoc Presentation of Research

12:30–13:30  Lunch





III. DATASETS, INFRASTRUCTURE AND INTEROPERABILITY

13:30–15:30 **Citation Extractor and Classifier: Pipeline and Datasets**

Angelo Di Iorio, Matteo Guenci, Marta Soricetti, Silvio Peroni, Lorenzo Paolini*, Ivan Heibi (University of Bologna) (*presenting)*


Open Citation Data for Educational Research

Tamara Heck, Christoph Schindler, Verena Weimer, Philipp Mayr, Ahsan Shahid (DIPF / GESIS)

Datasets in the Legal Theory Knowledge Graph Project

Christian Boulanger, Andreas Wagner (mpilhlt)

Interoperability Roundtable: Open Discussion on Data Models and Data Formats

15:30-16:00  Coffee

IV. TOOLS, WORKFLOWS AND PIPELINES

16:00–17:30 **LLM-Based Knowledge Graph Extraction Pipeline**

Raphael Schlattmann (TU Berlin), Malte Vogl (mpigea), Aleksandra Kaye (mpigea)


Training the Grobid Reference Extraction Models

Luca Foppiano (ScienziaLAB)

Annotation Tools for Machine Learning: PDF-TEI Editor (for LLaMora & Grobid), Prodigy, TEI-Publisher

Christian Boulanger, Andreas Wagner (mpilhlt)

17:30–18:30 **Takeaways, Way Forward, Closing**

19:00  Dinner (self-paid), Restaurant [Zur Stalburg](#), Glauburgstraße 80, 60318 Frankfurt am Main

Registration is possible at <https://plan.events.mpg.de/e/refextract25>. A link for online attendance will be sent to registered participants before the event.

The workshop series homepage is at: <https://mpilhlt.github.io/reference-extraction/>.

Contact:

Christian Boulanger <boulanger@lhlt.mpg.de>

Andreas Wagner <wagner@lhlt.mpg.de>

