LM4DH 2025

Proceedings of
The First Workshop on Natural Language Processing and Language Models for Digital Humanities

associated with

The 15th International Conference on Recent Advances in Natural Language Processing RANLP'2025

Edited by Isuri Nanomi Arachchige, Francesca Frontini, Ruslan Mitkov and Paul Rayson

11 September, 2025
Varna, Bulgaria

The First Workshop on Natural Language Processing
and Language Models for Digital Humanities
Associated with the International Conference
Recent Advances in Natural Language Processing
RANLP'2025

PROCEEDINGS

Varna, Bulgaria
11 September 2025

Online ISBN 978-954-452-106-6

Designed by INCOMA Ltd.
Shoumen, BULGARIA

Welcome to the LM4DH workshop

Digital Humanities has emerged as an interdisciplinary field of research, serving as an intersection of computer science with many other fields such as linguistics, social sciences, history, psychology, etc. With the development of Large Language Models (LLMs), state-of-the-art Natural Language Processing (NLP) tasks such entity recognition, text summarisation, diachronic analysis, and sentiment modelling have been significantly enhanced, offering powerful tools to analyse and interpret complex historical and cultural data. These advancements provide powerful tools for analysing and interpreting intricate historical, cultural, and social data, enabling researchers to identify patterns, extract meaningful relationships, and generate interpretations at unprecedented scale and precision.

Language Models for Digital Humanities (LM4DH) 2025 convened a collaborative platform for researchers, practitioners, and students to explore, critique, and advance AI-driven methodologies. We aimed to share technical innovations while fostering a community dedicated to ethically grounded and socially meaningful applications of LLMs.

We received 23 high-quality submissions for the LM4DH 2025 workshop, spanning a diverse range of topics at the intersection of language models and Digital Humanities: including computational linguistics, historical document processing, music augmentation, rhetorical analysis, sociolinguistic forecasting, ancient language parsing, and mental health text classification. Following a rigorous peer-review process, 18 papers were accepted for presentation and publication in the workshop proceedings.

The workshop featured two distinguished keynote speakers who offered valuable insights at the intersection of computational linguistics and digital humanities. Dr. Alessio Miaschi from the ItaliaNLP Lab at the Istituto di Linguistica Computazionale (CNR-ILC), Pisa, delivered a talk titled "From LLM Evaluation to Digital Social Reading," in which he examined the interpretability and evolution of neural language models and their growing relevance to linguistic research, outlining key scenarios where NLP can enrich humanistic modes of reading and interpretation. Complementing this, Professor Paul Rayson from Lancaster University showcased the practical applications of digital humanities in geospatial narrative processing, demonstrating how computational tools can map and contextualise historical testimonies across time and space. He further reflected on the future trajectory of the field, underscoring the indispensable role of interdisciplinary collaboration in ensuring the methodological rigour, innovation, and societal impact of digital humanities projects.

The success of LM4DH 2025 would not have been possible without the generous contributions of many exceptional individuals who supported this initiative. First and foremost, we extend our deepest gratitude to the authors who submitted their innovative work, helping to advance the vital intersection of language models and Digital Humanities. We are equally indebted to the members of the Program Committee, whose thoughtful engagement, timely reviews, and incisive feedback were instrumental in shaping the workshop's scholarly quality. Their dedication not only elevated the rigor of accepted submissions but also ensured the program reflected the highest standards of academic excellence and interdisciplinary innovation. Together, they have not only documented the state of the art but have helped define its future.

Organisers of LM4DH 2025

Organising Committee

Isuri Anuradha, Lancaster University, UK
Deshan Sumanathilake, Swansea University, UK
Francesca Frontini, CNR , Italy
Paul Rayson, Lancaster University, UK
Ruslan Mitkov, Lancaster University, UK

Programme Committee

Maram Alharbi, Lancaster University, UK
Salmane Chafik, Mohammed VI Polytechnic University, Morocco
Ignatius Ezeani, Lancaster University, UK
Safia Kanwal, Swansea University, UK
Chamila Liyanage, University of Colombo, Sri Lanka
Laura Noriega Santianez, University of Malaga, Spain
Yael Netzer, Hebrew University, Israel
Saadh Jawwadh, Informatics Institute of Technology, Sri Lanka
Gayanath Chandrasena, University of Helsinki, Finland
Marie Escribe, Universitat Politecnica de Valencia (UPV), Spain
Federico Boschetti, CNR Istituto di, Linguistica Computazionale A. Zampolli, Italy
Voula Giouli, Aristotle University of Thessaloniki, Greece

Table of Contents