1st Call for Papers

The 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)
Bangkok, Thailand
Thursday, August 15 2024 (co-located with ACL 2024)
http://ml4al.com

Submission deadline: May 17th, 2024 11:59pm, UTC-12 (anywhere on Earth)

DESCRIPTION

Ancient languages preserve the cultures and histories of the past. However, their study is fraught with difficulties, and experts must tackle a range of challenging text-based tasks, from deciphering lost languages to restoring damaged inscriptions, to determining the authorship of works of literature. Technological aids have long supported the study of ancient texts, but in recent years advances in Artificial Intelligence (AI) and Machine Learning (ML) have enabled analyses on Ancient Languages on an unprecedented scale and in unparalleled detail. The ML4AL Workshop aims to inspire and support research momentum in the emerging field of ML for the study of ancient texts.

The written evidence of the Ancient World is multifaceted and expansive. We invite contributions tackling texts from the diverse corners of the globe, in any language, script or medium. We establish a chronological scope from the inception of writing systems in ancient Mesopotamia and Egypt (3400 BCE) to the late first millennium CE. Encompassing such a vast and fertile remit for ML applications, the ML4AL Workshop is designed to facilitate and invigorate the ongoing collaborative momentum between ML and the Humanities, to foster a deeper understanding of our past. Indeed, ancient languages fall under the category of low-resource languages due to the scarcity of available linguistic data for modern analysis. These languages, therefore, offer a compelling case study for ML: their limited textual material, socio-cultural intricacies, evolving forms, and diverse transmission histories pose a significant challenge to conventional models.

We welcome contributions on topics related to, but not limited to:
- Digitization: bringing textual sources to a high-quality machine-readable format (e.g., through HTR).
- Restoration: recovering missing text and reassembling fragmented written artefacts.
- Attribution: contextualising a document within its original geographical, chronological and authorial setting.
- Linguistic analysis: involving tasks such as POS tagging, text parsing, segmentation, representation learning, semantics, sentiment, language identification.
- Textual criticism: the process of reconstructing a text's philological tradition of textual transmission, including the tasks of stemmatology and intertextuality.
- Translation and decipherment: which aim to make a text's language comprehensible and interpretable to modern-day researchers.

We particularly welcome submissions which tackle low-data, underrepresented, non-Western ancient languages, and we encourage researchers and practitioners from

diverse backgrounds, working on ancient languages, irrespective of their gender, ethnicity, nationality, or academic affiliations, including fellows tackling low-underrepresented and non-Western centric ancient languages.

SUBMISSION INFORMATION

We welcome long (8 page) and short (4 page) paper submissions, in PDF format, made through OpenReview or ARR. Accepted regular workshop papers will be included in the workshop proceedings, but non-archival submissions are also welcome:

Regular workshop papers: Both long (8 pages) and short (4 pages) papers may have unlimited pages for references and up to 100 MB of supplementary materials (separately). Authors are strongly encouraged to submit their code for reproducibility. In the camera-ready version, one additional page of content will be given to address the comments received by the reviewers. All submissions should be completely anonymous to allow a double-blind review process and the papers should follow the ACL template style. Each paper is expected to be reviewed by at least three reviewers. Selected accepted papers will be presented orally and the rest as posters.
Non-archival submissions: Papers on relevant topics that have appeared or might appear in other venues (workshops, conferences, journals) are also welcome, which can be presented at the workshop but will not be included in the workshop proceedings.

Already published contributions (excluding preprints) cannot be accepted. Papers being submitted both to ML4AL and another venue must note on the title page the other conference/workshop and state on the title page that if the authors choose to present their paper at ML4AL (upon acceptance), then the paper will be withdrawn from other conferences and workshops. All submitted manuscripts should be fully anonymous (please avoid self-references) and must include a dedicated "Limitations" section, which will not count toward the page limit. Supplementary material (e.g., code, data, audio/visual material, etc.) is suggested to be uploaded on a repository, anonymously, and linked to the paper.

ORGANIZING COMMITTEE

Dr John Pavlopoulos, Athens University of Economics and Business, Greece
Dr Thea Sommerschield, University of Nottingham, UK
Dr Yannis Assael, Google DeepMind, UK
Dr Shai Gordin, Ariel University, Israel
Prof. Kyunghyun Cho, NYU, CIFAR, Genentech, USA
Prof. Marco Passarotti, Università Cattolica del Sacro Cuore, Italy
Dr Rachele Sprugnoli, Università di Parma, Italy
Dr Yudong Liu, Western Washington University, USA
Dr Bin Li, Nanjing Normal University, China
Dr Adam Anderson, UC Berkeley, USA

Contact the organizers at: ml4al.organizers@gmail.com

IMPORTANT DATES

- Paper submission deadline: May 17, 2024
- Notification of acceptance: June 17, 2024
- Camera-ready paper due: July 1, 2024
- Workshop: August 15, 2024

All deadlines are 11:59 pm UTC -12h ("anywhere on Earth")

PROGRAM COMMITTEE

Masayuki Asahara; John Bodel; Gregory Crane; Katrien De Graef; Sanhong Deng; Mark Depauw; Hanne Eckhoff; Margherita Fantoli; Minxuan Feng; Ethan Fetaya; Federica Gamba; Laura Hawkins; Chul Heo; Petra Heřmánková; Marietta Horster; Renfen Hu; Kyle Johnson; Alek Keersmaekers; Ussen Kimanuka; Thomas Koentges; Els Lefever; Chaya Liebeskind; Eliese-Sophia Lincke; Chao-Lin Liu; Liu Liu; Congjun Long; Jiaming Luo; Massimo Maiocchi; Isabelle Marthot-Santaniello; Barbara McGillivray; M. Willis Monroe; Alex Mullen; Chiara Palladino; Chanjun Park, Upstage; Edoardo M. Ponti; Mladen Popovic; Jonathan Prag; Avital Romach; Edgar Roman-Rangel; Matteo Romanello; Brent Seales; Andrew Senior; Si Shen; Barak Sober; Richard Sproat; Gabriel Stanovsky; Vanessa Stefanak; Silvia Stopponi; Qi Su; Matthew I. Swindall; Xuri Tang; Charlotte Tupman; Dongbo Wang; Haneul Yoo; Chongsheng Zhang