LaTeCH-CLfL 2024

The 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature

Proceedings of the Workshop

March 22, 2024

The LaTeCH-CLfL organizers gratefully acknowledge the support from the following sponsors.

IDeaL
SFB 1102

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-069-1

Introduction

Welcome to the 2024 edition of LaTeCH-CLfL! Whether you are coming back or joining us for the first time, we are delighted to have you here. This workshop, with a history of nearly two decades, continues to serve as home for a wide spectrum of discussions. This year is no exception, with a lineup of topics that span the intersection of language technology, computational linguistics and the broadly conceived humanities.

This year, we keep studying literature through the computational lens, exploring the complex interplay of literary devices, individual style and the markers of narrative structures. This group of works includes papers that focus on quantitative analyses and experiments to study literary texts or verify literary hypotheses, such as automatic recognition of knowledge transfer in German drama, authorship verification models for quotation attribution, and sentiment analysis in a low resource setting. These papers demonstrate the application of computational methods in the analysis of literary texts and authors' styles.

Computational methods are increasingly prominent in the study of history, reporting new dimensions of large-scale phenomena and their development across centuries. Papers in this category apply computational techniques to historical and cultural sources, including handwritten text recognition in marginalia, named entity recognition in historical texts, topic modelling to explore historical portrayals, the analysis of diachronic scientific writing and the study of parliamentary debates. This research showcases the intersection of digital humanities with computational linguistics to uncover insights from historical and cultural texts.

This year, we highlight the latest innovations in language processing tools and digital resources to cover both diachronic and synchronic differences. This group of papers focuses on the development and evaluation of computational tools and resources, such as coreference resolution in a corpus of long documents, entity linking in digital content, PoS tagging of Latin texts through GPT and the OCR correction of historical text transcripts. These contributions are essential for advancing methodologies in text analysis and improving the accuracy of computational models.

Last but not least, the workshop encompasses papers dealing with sociopolitical and cultural text analysis. Included here are papers that analyze contemporary themes and emergent phenomena through computational methods, such as the dehumanization of Ukrainians on social media, metaphorical framing in media, topic modelling of newspapers, and the effects of the Plain English Movement on legal and scientific articles. This research highlights the role of computational linguistics in understanding contemporary events and how they are discussed in the public forum.

As you can see, there is something for everyone (all things considered) but do keep an open mind and read all papers, if you have the time. You will be glad you did.

Do not forget to visit our Web site HERE – and check out past workshops too.

It goes without saying that whatever success our workshop enjoys is due to the authors (thank you for staying with us or for trusting us the first time), and without question to the reviewers. A special shout-out to our wonderful program committee!

Yuri, Stefania, Anna, Stan

Organizing Committee

Organizers

Yuri Bizzoni, Aarhus University
Stefania Degaetano-Ortlieb, Saarland University
Anna Kazantseva, National Research Council Canada
Stan Szpakowicz, EECS, University of Ottawa

Program Committee

Program Committee

Diego Alves, Saarland University
Melanie Andresen, Universität Stuttgart
Jinyeong Bak, Sungkyunkwan University
Paul Buitelaar, University of Galway
Miriam Butt, University of Konstanz
Florian Cafiero, Ecole nationale des chartes - PSL
Jacob Eisenstein, Google
Anna Feldman, Montclair State University
Mark Finlayson, FIU
Stefan Fischer, Universität des Saarlandes
Antske Fokkens, VU Amsterdam
Francesca Frontini, Istituto di Linguistica Computazionale A. Zampolli"- ILC Consiglio Nazionale delle Ricerche - CNR
Udo Hahn, Friedrich-Schiller-Universitaet Jena
Serge Heiden, ENS de Lyon
Rebecca Hicke, Cornell University
Labiba Jahan, Southern Methodist University
Fotis Jannidis, Universität Würzburg
Dimitrios Kokkinakis, University of Gothenburg
Stasinos Konstantopoulos, NCSR Demokritos
Marie-Pauline Krielke, Saarland University
Maria Kunilovskaya, Saarland University
John Ladd, Washington and Jefferson College
John Lee, City University of Hong Kong
Chaya Liebeskind, Jerusalem College of Technology , Lev Academic Center
Thomas Lippincott, Johns Hopkins University
Barbara Mcgillivray, King's College London
David Mimno, Cornell University
Vivi Nastase, University of Geneva
Borja Navarro-Colorado, University of Alicante
Pierre Nugues, Lund University
Petya Osenova, Sofia University St. Kl. Ohridskiand IICT-BAS
Janis Pagel, Department of Digital Humanities, University of Cologne
Andrew Piper, McGill University
Petr Plecháč, Institute of Czech Literature CAS
Thierry Poibeau, LATTICE (CNRS and ENS/PSL)
Jelena Prokic, Leiden University
Georg Rehm, DFKI
Nils Reiter, University of Cologne
Martin Reynaert, ILLC - Universiteit van Amsterdam / DCA - Tilburg University
Pablo Ruiz Fabo, LiLPa, Université de Strasbourg
Marijn Schraagen, Utrecht University
Artjoms Šeļa, Institute of Polish Language (PAN)
Daniil Skorinkin, Higher School of Economics (Moscow, Russia)
Pia Sommerauer, Vrije Universiteit Amsterdam
Elke Teich, Universität des Saarlandes

Laure Thompson, University of Massachusetts Amherst
Ulrich Tiedau, University College London
Rob Voigt, Northwestern University
Albin Zehe, University of Wuerzburg
Heike Zinsmeister, Universitaet Hamburg

Additional Reviewers

Sam Backer, John Hopkins University
Pascale Feldkamp, Center for Humanities Computing, Aarhus University
Jörg Knappen, Saarland University
Ida-Marie Lassen, Center for Humanities Computing, Aarhus University
Craig Messner, John Hopkins University
Hale Sirin, John Hopkins University

Keynote Talk: The Impresso Project's Approach to Historical Media Analysis

Marten Düring and Maud Ehrmann
Luxembourg Centre for Contemporary and Digital History (C2DH) // Digital Humanities Laboratory of the Ecole Polytechnique Fédérale de Lausanne

Abstract: In the impresso project, a team of computational linguists, historians and designers strive to enable data-driven analyses of historical media across time, institutional silos, media types and languages. To this end we compile a corpus of historical newspaper and radio collections with the help of Western European partners, enrich it using text mining techniques and develop user interfaces for their exploration and computational analysis.

Bio:

Marten has a robust background at the crossroads of cultural history, digital humanities, and computational methods. He has a rich academic trajectory, including a PhD in Contemporary History from the University of Mainz, with his dissertation focusing on the emergence of covert networks during World War II. Marten has contributed his expertise to various institutions, such as Radboud University and the University of North Carolina at Chapel Hill, and has been integral to digital history initiatives at C2DH since 2016.
At C2DH, Marten serves as a principal investigator for the impresso project, which seeks to revolutionize the way historical newspaper and radio collections are accessed and analyzed, transcending barriers of language and national borders. His role encompasses coordinating interface development and steering digital history research within the team. Additionally, Marten is a founding editor of the Journal of Historical Network Research and leads the coordination of the Historical Network Research Community. His commitment to the field is also evident through his involvement in the Hands-on History lecture series and his proactive engagement in support activities for Ukrainian scholars by the C2DH center following the 2021 crisis.
Marten's work not only reflects his dedication to enhancing the tools available for historical research but also underscores the potential of interdisciplinary approaches that meld historical inquiry with technological innovation. His presence as a speaker is a testament to his leadership in shaping the future of digital history.

Maud is a research scientist and lecturer at the Digital Humanities Laboratory of the Ecole Polytechnique Federale de Lausanne. She holds a PhD in Computational Linguistics from the Paris Diderot University (Paris 7) and has been engaged in a large number of scientific projects centred on information extraction and text analysis, both for present-time and historical documents. Before joining the DHLAB, she worked at the Linguistics Computing Laboratory at the Sapienza University of Rome where she worked on the BabelNet resource and contributed to the LIDER project (2013-2014). Prior to that, she worked at the European Commission's Joint Research Centre in Ispra, Italy, as member of the OPTIMA unit (now Text and Data mining unit) which develops innovative and application-oriented solutions (Europe Media Monitor) for retrieving and extracting information from the Internet with a focus on high multilinguality (2009-2013). Previously, she worked at the Xerox Europe Research Centre in Grenoble, France (now Naver Labs Europe) in the Parsing and Semantics unit, first as PhD candidate supported by a CIFRE grant (2005-2008), then as a post-doctoral researcher (2008-2009). There, her research focused mainly on the automatic processing and fine-grained analysis of entities of interest, specifically named entities and temporal expressions.

Table of Contents