HistoInformatics2021: The 6 th International Workshop on Computational History

Yasunobu Sumikawa, Ryohei Ikejiri, Antoine Doucet, Eva Pfanzelter, Mohammed Hasanuzzaman, Gaël Dias, Ian Milligan, Adam Jatowt

HistoInformatics2021: The 6th International Workshop on Computational History

Yasunobu Sumikawa
Dept. of Computer Science
Takushoku University
Tokyo, Japan
ysumikaw@cs.takushoku-u.ac.jp

Ryohei Ikejiri
Interfaculty Initiative in Information Studies
The University of Tokyo
Tokyo, Japan
ikejiri@iii.u-tokyo.ac.jp

Antoine Doucet
L3i laboratory
University of La Rochelle
La Rochelle, France
antoine.doucet@univ-lr.fr

Eva Pfanzelter
Department of Contemporary History
University of Innsbruck
Innsbruck, Austria
eva.pfanzelter@uibk.ac.at

Mohammed Hasanuzzaman
ADAPT Centre, Munster Technological University
Cork, Ireland
mohammed.hasanuzzaman@adaptcentre.ie

Gaël Dias
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC
Caen, France
gael.dias@unicaen.fr

Ian Milligan
Dept. of History, Faculty of Arts, University of Waterloo
Waterloo, Canada
i2milligan@uwaterloo.ca

Adam Jatowt
Institute for Informatics
University of Innsbruck
Innsbruck, Austria
adam.jatowt@uibk.ac.at

***Abstract*— We propose the HistoInformatics2021 (The 6th International Workshop on Computational History) workshop to be held in conjunction with the JCDL2021 conference. This would be the 6th installment of the workshop series devoted to the interaction between Computer Science and History. This interdisciplinary initiative is a response to the growing popularity of Digital Humanities, particularly in historical research, and an increased tendency to apply algorithms and computer techniques for fostering and facilitating new research methods and tools in the Humanities.**

Keywords—computational history, digital history, digital humanities

I. Objectives of the Workshop

The proposed workshop will cover a wide array of subjects related to the growing field of HistoInformatics. Our objective is to provide for the two different research communities: Computer Science (CS) and History (H) a place to meet and exchange ideas and to facilitate discussion. We hope the workshop will result in a survey of current problems and potential solutions, with a particular focus on exploring opportunities for collaboration and interaction of researchers working on various subareas within Computer Science and History.

II. Intellectual Background

History as a representation of the past has many functions. It helps to create meaning, coherence, and orientation for individuals and collectives and some believe that it helps settle the foundations of our nations, our identities and our memories, to name a few. As such, it is one of the fundamental subjects taught from elementary schools onwards. Traditionally historical research is based on manual investigation of preserved records and artifacts to provide a reliable account of the past. Alongside this hermeneutic approach some historians have translated primary sources into data and used statistics to analyze them. More recently, the field of Digital History has received attention with its aim to explore historical resources in a variety of ways through the usage of automatic methods.

Nowadays, due to the constantly growing number of repositories which provide access to digitized documents, the field of History can greatly benefit from the advances of Computer and Information sciences which consist of processing, organizing and making sense of data and information. As such, new Computer Science techniques can be applied to help verify and validate historical assumptions based on text analytics, image interpretation or comparing multiple perspectives. Hence, Digital History has now entered a new era that we call HistoInformatics, analogous to BioInformatics and ChemoInformatics which have respectively proposed new research trends in Biology and Chemistry. Many subareas of Computer science are relevant and have potential to advance historical studies. These include Information Retrieval, Information Extraction, Text Analytics, Natural Language Processing, Artificial Intelligence (Machine Learning, Knowledge Representation), Image Processing, and others.

III. Identification of expected audience

The main topic of the workshop is that of supporting historical research and analysis through the application of Computer Science theories or technologies, analyzing and making use of historical texts, and simulating past courses of actions, analyzing collective memories, visualizing historical data and providing efficient access to the large wealth of historical knowledge. The workshop will be open to any participants who are interested in the convergence of history and computer sciences. Based on the number of attendants in the previous editions of the workshop (20-40) we expect the participation in HistoInformatics2021 to be similar if not better. We expect the audience to consist of DL researchers interested in digital humanities projects as well as computer science and history science researchers who wish to learn about digital library topics and technologies.

IV. Biographies of organizers

Yasunobu Sumikawa is an Assistant Professor at the department of computer science, Takushoku University, Japan. His research interests lie on information retrieval, computational history, and history learning. He is currently involved with several research projects about causality mining for enhancing historical analogy funded by the Grants-in-Aid for Scientific Research. He is currently leading a machine learning research group.

Ryohei Ikejiri is a Lecturer (Project Assistant Professor) in the Interfaculty Initiative in Information Studies, the University of Tokyo, Japan. His research interests include history education, transfer of historical learning and educational technology. He is the recipient of five thesis awards for his educational technology papers on history learning. He is currently conducting some research projects about developing digital learning materials including AI technology and game mechanics to support historical transfer, historical inquiry and historical thinking. He has been actively involved in a Public History Research Group in Japan.

Antoine Doucet is a Full Professor in computer science in the university of La Rochelle. His main research interests lie in the fields of information retrieval (structured and semi-structured) and natural language processing (NLP). The central focus of his work is on the development of methods that scale to very large document collections and that do not require prior knowledge of the data: in particular, techniques that function for documents written in any language and that are robust to noise (from imperfect OCR or speech-to-text for instance). Since 2018, he is the coordinator of the NewsEye project (A Digital Investigator for Historical Newspapers) funded by the Horizon 2020 programme. Additionally, he leads the digital image and contents research group since 2016, which currently consists of about 50 researchers.

Eva Pfanzelter is an Associate Professor at the Department of Contemporary History at the University of Innsbruck (Austria). Her main focus is on digitizing historical material and using historical online content for research and teaching in contemporary history, computational methods, and digital history in theory and practice. For her upcoming publication (summer 2021) on "Digital Holocaust. Negotiations of Genocide between Public History, Politics of History and Commerce" she combined big data approaches with qualitative analysis to come to terms with data from internet archives, wikipedia and social networks. Within the EU-funded project "NewsEye. A Digital Investigator for Historical Newspapers" she is leading the Digital Humanities group.

Mohammed Hasanuzzaman is a Lecturer (Assistant Professor) at the department of computer science, Munster Technological University (Cork Campus), Ireland. He is also a Funded Investigator of the €100 million Irish Government-funded ADAPT Centre for Global Digital Content (http://adapt-centre.ie/). His research interests and activities over the past 8 years have been in Natural Language Processing (NLP), especially Temporal Information Extraction, Machine Translation, Psycholinguistics, Social Media Analytics, e-health and Machine Learning (Deep Learning) applications to NLP. He has published more than 40 papers in a range of highly respected conferences and journals. He is currently leading (as PI/Co-PI) several projects such as Watching the risk factors (WARIFA): Artificial intelligence and the prevention of chronic conditions, STop Obesity Platform, STEM-in Action; Open Educational Resources for Teachers funded by the Horizon 2020 programme. He is previously involved in the organization of the HistoInformatics workshop series. He had co-edited the special issue on HistoInformatics: Computational Approaches to History in the Journal of Data Mining and Digital Humanities.

Gaël Dias is a Full Professor at the Department of Computer Science of the University of Caen Normandie (France) and develops his research activities at the CNRS GREYC UMR 6072 Laboratory. His research interests include lexical semantics, information digestion, Web accessibility for visually impaired people, multimodal NLP for mental health, and computational history. He is co-director of the GREYC Lab and has published more than 140 papers in top international conferences and journals (AAAI, ACL, EACL, ECAI, SIGIR, WWW, etc.). He has been co-organizing 4 out of 5 of the HistoInformatics workshop series. He co-organized 13 workshops overall (6 at EPIA, 1 at ACL, 1 at SIGIR, 1 at CIKM, 1 at LREC, 1 at SocInfo, 1 at TPDL) and two main conferences (TALN 2015, EPIA 2005). He co-organized the special issue on HistoInformatics: Computational Approaches to History at the Journal of Data Mining and Digital Humanities.

Ian Milligan is an Associate Professor at the Department of History at the University of Waterloo (Canada). His primary research focus is on how historians can use web archives, as well as the impact of digital sources on historical practice more generally. He is Principal Investigator of the Andrew W. Mellon Foundation-funded Archives Unleashed Project, which works with the Internet Archive to develop better access tools for their collections. He served as a PC co-chair of JCDL 2017, and has also organized a series of 7 Archives Unleashed datathons, as well as having served as a workshop and poster co-chair at various JCDL conferences.

Adam Jatowt is a Full Professor at the Institute of Informatics at University of Innsbruck, Austria. His research interests include computational history, temporal information retrieval from text, web archive search and mining, and information comprehensibility. He served as a PC co-chair of IPRES2011, ICADL2014, JCDL2017, ICADL2019 and a general chair of TPDL2019 and ICADL2020. He also serves as a Steering Committee Chair of ICADL from 2019. He is the founder as well as was the co-organizer of every HistoInformatics workshops. Adam has in total organized 23 international workshops including 7 at WebConf (WWW), 4 at ECIR, 2 at ACL, 2 at CIKM, 2 at SocInfo, 1 at DH, 1 at TPDL and 1 at IJCAI conferences.

V. Anticipated outcomes

We hope to attract reports on complete and ongoing research and implementations as well as seek position papers that will survey the existing solutions and introduce possibilities for further work. Accepted papers will be published either as full papers up to 8 pages or as short synopsis of length up to 4 pages as CEUR-WS proceedings. We are also considering running a special issue based on selected papers at the Journal of Data Mining and Digital Humanities (JDMDH). We have already received the consent from the journal's editors. We plan to invite 1-2 prominent keynote speakers who work in the digital history area. Finally, the workshop could result in novel collaborations between historians, humanities, and computer science researchers. We have observed such spontaneous interactions in the past editions of the workshop.