GLOBAL STUDIES WORKSHOP ON DIGITAL INFORMATION RETRIEVAL FROM LARGE COLLECTIONS OF DATA

ORGANISER: Global Studies Research Programme

CONTACT: Mette Thunø (mettethunoe@cas.au.dk)

TIME: Wednesday Sep. 23, 2020, 9.00-12.30

LOCATION: 1467-116 (on-campus workshop with one metre distance between participants)

REGISTRATION: please answer 'accept' to digital calendar invitation to be sent to your mailbox

Global Studies Research Programme has invited Center for Humanities Computing, CAS, AU to organise a workshop on new methods and tools to work with large collections of text, but also visual data. Two colleagues will also present their own research experiences with large data collections and we will discuss the implications of these new research methods for our research.

Digitization and automation are increasingly offering us new tools and research methods that allow the extraction of useful information from large collections of text and visual data from e.g. media, public institutions. The majority of these tools and methods favour data from the Western world encoded in English. We are, however, seeing an increase in language technology tools also for languages covered at Global Studies.

In this workshop, Center for Humanities Computing presents an excerpt of digital technologies for accessing 'non-Western' text and visual data to extract meaningful patterns that can be utilized both in qualitative and quantitative analysis. We will introduce three sets of approaches to information extraction: 1) ngram-based approaches that capture how often words or sequences of words occur in a collection of texts, 2) vector-based approaches that can be used to model concepts and discursive structures, and 3) sentiment and emotion detection that identify affective properties of texts and images.

PROGRAMME:

9.00-9.10:	Mette Thunø: Welcome
9.10-9.20:	LIU Jun, Associate Professor, University of Copenhagen: "Bias, incompleteness, and distortion: The challenge to text-based computational methods in censorship studies"
9.20-9.30:	Lisanne Wilken, Associate Professor, Aarhus University: "Big and Thick: Computational and ethnographic approaches to analysis of multilingual & multimodal twitter 'conversation' during the Eurovision Song Contest finals in recent years"
9.30-9.45:	Q&A
9.45-10.45:	Kenneth Enevoldsen, Peter B. Vahlstrup, Ross D. Kristensen-McLachlan and Kristoffer L. Nielbo from Center for Humanities Computing
10.45-11.00:	Coffee and tea break
11.00-12.00:	Center for Humanities Computing continues
12.00-12.30:	Discussions

AARHUS
UNIVERSITY