AESPEN: Automated Extraction of Socio-political Events from News @ LREC 2020

Call for Papers

Automatic construction of event databases has long been a challenge for the NLP community in terms of algorithmic approaches and language resources. At the same time, social and political scientists have been working on creating socio-political event databases for their studies for decades. The need for collecting data for socio-political science studies has been satisfied by manual, semi-automatic and automatic approaches. However, the results yielded by these approaches to date are either not of sufficient quality or require tremendous effort to be replicated on new data. On the one hand, manual or semi-automatic methods require high quality human effort; on the other hand, state-of-the-art event detection systems are not accurate to the point that their output may be directly usable, without human moderation. Finally, the natural language processing (NLP) community has not achieved a consensus on the treatment of events both in terms of task definition and appropriate techniques for their detection.

Given the aforementioned limitations, there is an increasing tendency to rely on machine learning (ML) and NLP methods to deal better with the vast amount and variety of data to be collected. The aim of this workshop is to inspire the emergence of innovative technological and scientific solutions in the field of event detection and event metadata extraction from news, as well as the development of evaluation metrics for event recognition. At the same time the workshop will aim at triggering a deeper understanding of the usability of socio-political event datasets.

References:


Motivation and Topics of interest

Automating political event collection requires availability of gold-standard corpora that can be used for system development and evaluation. Moreover, automated tool performances need to be reproducible and comparable. Although tremendous effort is being spent on creating socio-political event databases such as ACLED, GDELT, MMAD, and ICEWS, there has not been much progress in harmonization across event schemas and tasks. This limitation causes the definition of the events and automated event information collection tool performances to be restricted to single projects. Consequently, the lack of comparable and reproducible settings hinders progress on this task.

We invite contributions from researchers in NLP, ML and AI involved in automated event data collection, as well as researchers in Social and Political Sciences, Conflict Analysis and Peace studies, who make use of this kind of data for their analytical work. Our goal is to enable the emergence of innovative NLP/IE solutions that can help dealing with the current stream of information, manage the risks of information overload, identify different sources and perspectives, and provide unitary and intelligible representations of the larger and long-term storylines behind news articles.

Our workshop will provide a venue for discussing about the creation and facilitation of language resources in the social and political sciences domain. Social and political scientists will be interested in reporting and discussing their desiderata from automated tools in comparison to their traditional coding approaches. Computational linguistics and machine learning practitioners and researchers will benefit from being challenged by real-world use cases, in terms of event data extraction, representation and aggregation.

We invite work on all aspects of automated coding of socio-political events from mono- or multi-lingual news sources. This includes (but is not limited to) the following topics:

- Event metadata extraction
- Source Bias mitigation
- Event Data Schema and representation
- Event information duplication detection
- Extracting events beyond a sentence in a document
- Training Data collection/annotation processes
- Event coreference (in- and cross-document)
Sub-event and event subset relations
Event dataset evaluation metrics
Event datasets quality assessments
Defining, populating and facilitating event ontologies
Automated Tools for relevant tasks
Understanding the limits that are introduced by copyright rules
Ethical concerns and ethical design

Submissions

This call solicits full papers reporting original and unpublished research on the topics listed above. The papers should emphasize obtained results rather than intended work and should indicate clearly the state of completion of the reported results. Submissions should be between 4 and 8 pages in total.

Authors are also invited to submit short papers not exceeding 4 pages (plus two additional pages for references). Short papers should describe:

- a small, focused contribution;
- work in progress;
- a negative result;
- a position paper.

Proceedings should be submitted on the START system, in compliance with the style sheet adopted for the LREC Proceedings (to be found here: https://lrec2020.lrec-conf.org/en/submission2020/authors-kit/)

The reviewing process will be double blind and papers should not include the authors’ names and affiliations. Each submission will be reviewed by at least three members of the program committee. If you do include any author names on the title page, your submission will be automatically rejected. In the body of your submission, you should eliminate all direct references to your own previous work.

Workshop Proceedings will be published on the LREC 2020 website.

Identify, Describe, and Share your LRs!

Describing your LRs in the LRE Map is now a normal practice in the submission procedure of LREC (introduced in 2010 and adopted by other conferences). To continue the efforts initiated at LREC 2014 about “Sharing LRs” (data, tools,
web-services, etc.), authors will have the possibility, when submitting a paper, to upload LRs in a special LREC repository. This effort of sharing LRs, linked to the LRE Map for their description, may become a new “regular” feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.

As scientific work requires accurate citations of referenced work so as to allow the community to understand the whole context and also replicate the experiments conducted by other researchers, LREC 2020 endorses the need to uniquely Identify LRs through the use of the International Standard Language Resource Number (ISLRN, www.islrn.org), a Persistent Unique Identifier to be assigned to each Language Resource. The assignment of ISLRNs to LRs cited in LREC papers will be offered at submission time.

**Shared Task**

We will organize a shared-task that will provide a setting that consists of data, task definition, and evaluation schema. Participants of this shared-task will have the possibility to report their results in the workshop after peer-review of their working notes. A session will be dedicated to discuss the results of the shared task during the workshop.

**Keynote Speaker**

The keynote speech will be delivered by Prof. Clionadh Raleigh.

**Title:** Too soon? The limitations of AI for event data

**Bio:**

Prof. Clionadh Raleigh is a professor of Political Geography focused on modern disorder and political elite networks in developing states. She is the director of the ACLED project which produces and analyzes real-time data on political violence and protest in the world’s most unstable states. Moreover, she is recipient of two European Research Council Grants.

**Important dates**

All dates are in 2020 and (23:59 GMT+1)

- January 14th: Announcing the shared task
- February 14th: Cut-off date for the shared task results
- February 22nd: Workshop paper Submission deadline
- March 1st: Submission deadline for the working notes of the shared task
• March 13th: Notification of acceptance
• April 2nd: Camera-ready deadline
• May 12th, 2020: The workshop date

Contact
Do not hesitate to contact ahurriyetoglu@ku.edu.tr or vanni.zavarella@ec.europa.eu for any question or comment.

Organizing Committee
Ali Hürriyetoğlu (Koc University)
Hristo Tanev (European Commission - Joint Research Center)
Erdem Yörük (Koc University and University of Oxford)
Vanni Zavarella (European Commission - Joint Research Center)

Programme Committee
Svetla Boycheva (Institute of Information and Communication Technologies, Bulgarian Academy of Sciences)
Fırat Durusan (Koc University)
Burak Gürel (Koc University)
Bernardo Magnini (Fondazione Bruno Kessler (FBK))
Osman Mutlu (Koc University)
Jakub Piskorski (Polish Academy of Sciences)
Lidia Pivovarova (University of Helsinki)
Benjamin J. Radford (UNC Charlotte)
Clionadh Raleigh (University of Sussex)
Parang Saraf (Virginia Tech)
Philip Schrodt (Parus Analytical Systems)
Manuela Speranza (Fondazione Bruno Kessler, Trento)
Çağrı Yoltar (Koc University)
Aline Villavicencio (The University of Sheffield)

Kalliopi Zervanou (Eindhoven University of Technology)