

TASK 2: Fine-grained classification of Socio-political Events

1. Task Definition

The objective of this generalized zero-shot learning task is to classify short text snippets reporting socio-political events with fine-grained event types using the Armed Conflict Location & Event Data Project (ACLED) event taxonomy, which consists of **25 event subtypes (seen classes)** pertaining to political violence, demonstrations (rioting and protesting) and selected non-violent, politically important events. The task is to label text snippets using ACLED types and potentially other types of similar events not covered directly by ACLED (**unseen classes**).

2. Training Data

For the training purposes one can use event-related textual data of any kind, where we recommend considering exploiting the short text snippets reporting events which are part of the large event dataset created by ACLED (<https://acleddata.com/>), and which can be obtained from ACLED using this web page: <https://acleddata.com/data-export-tool/> for academic purposes.

The text snippets stored in the Event Database of ACLED are labelled with one of the 25 sub-event types as specified in the ACLED code book at: https://acleddata.com/acleddatanew/wpcontent/uploads/dlm_uploads/2019/01/ACLED_Code_book_2019FINAL.docx.pdf

For the sake of training purposes one can exploit as an inspiration the techniques for text normalization and cleaning of ACLED data, and baseline classification models trained using ACLED data described in <https://www.aclweb.org/anthology/2020.coling-main.584/>.

3. Test data

Test data for this task will be **around 1,000 text snippets** from news, web pages reporting socio-political and security-related events, and artificially created event descriptions labelled using **both the ACLED event taxonomy** (seen classes) and some **event types not being part of ACLED** event typology (unseen classes).

4. Evaluation

The registered participants will be provided a single test data file, where each line consists of two tab-separated elements, i.e., an ID (integer), followed by a text snippet reporting an event. The system response files should have per event a line with the event ID and an event label separated by a tab. One text snippet is to be annotated with a single event type label.

The systems will be evaluated mainly using: **Precision**, **Recall**, and **Micro** and **Macro F-1** metrics, where the last two are the most important ones.