Coding Actions and Fatalities in Militarized Interstate Confrontations (MICs), 2002 to 2024

Summer 2025

Description

This project uses an existing corpus of newspaper articles to derive models that identify (1) whether a MIC occurred, (2) what type of militarized interaction was described, (3) whether fatalities occurred, and (4) the range of fatalities, if any. The newspaper articles have been retrieved based on Boolean search terms associated with militarized conflict. However, more than 95% of the articles are still false positives for identifying MICs.

Task

As part of the preliminary test to determine suitability for the project, you will be asked to determine whether one country's forces caused the death of one or more military personnel of another country's forces. We define countries as members of the international system of states, with country and membership dates listed here: state system list.

We have a preliminary corpus of newspaper articles for the years 2015 to 2023, and those articles can be accessed here: news articles. Use this corpus of articles to determine (1) the dates on which military forces were killed in combat, (2a and 2b) an approximate range (or precise figure) of the number of those deaths, and (3) the countries involved.

Deliverables

- (1) the dates, (2a and 2b) the range of fatalities [with 2a=2b if precise estimate], and (3) the countries involved.

- Script saved in the Github link you share that includes the process, model creation, training, and testing. The script should be able to be run from start to finish without user intervention.

- If any pre-trained models are used, add a link to a Google Drive that contains the pre-trained model(s).

- A copy of your curriculum vitae (resume) as an accessible link.

Please email the above information to Dr. Andrea Underhill (atunderhill@ua.edu), using the subject line **GSoC Coding Actions and Fatalities in Militarized Interstate Confrontations (MICs), 2002 to 2024**. Include in the body of the email links to the information listed above.

Metrics

Submissions will be judged according to accuracy and our testing on an additional corpus of newspaper articles from the same period.