

SEMEVAL-2018 TASK: *Counting Events and Participants within Highly Ambiguous Data covering a very Long Tail*

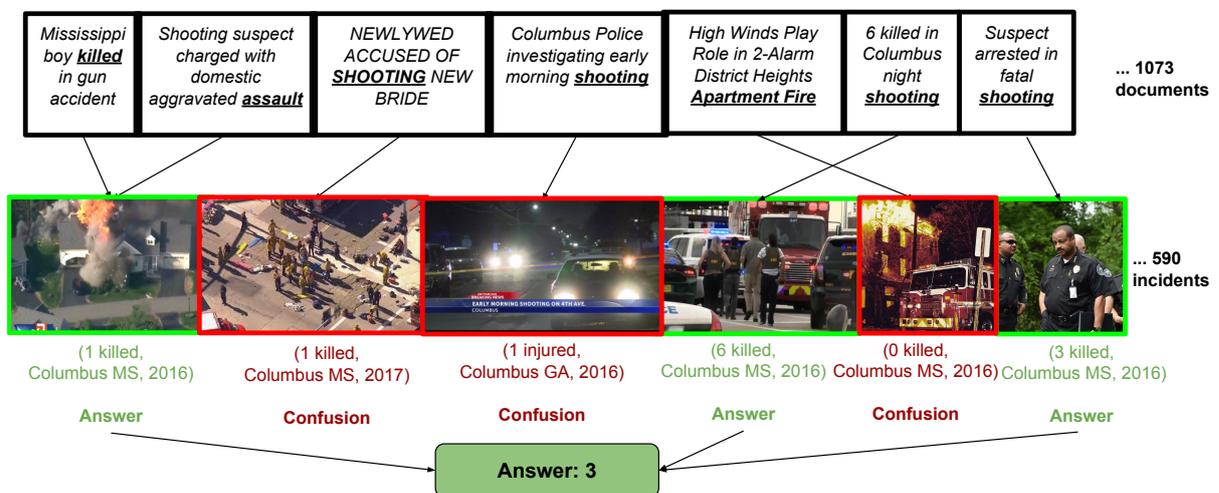


We are hosting a “referential quantification” task that requires systems to establish the meaning, reference and identity of events and participants in news articles. By “referential quantification”, we mean questions about the number of incidents of an event type (subtasks S1 and S2) or participants in roles (subtask S3), designed as three incrementally harder subtasks:

- Subtask S1** Event Questions with Answer=1
Which killing incident happened in 2014 in Columbus, OH?
- Subtask S2** Event Questions with Answer=any number
How many killing incidents happened in 2016 in Columbus, MS?
- Subtask S3** Participant Questions with Answer=any number
How many people were killed in 2016 in Columbus, MS?

Based on a set of questions and corresponding documents, the participant systems need to provide a numeric answer together with the supporting mentions of the answer events in the documents. A schematic example of the S2 challenge is given below:

Question: How many killing incidents happened in 2016 in Columbus, Mississippi?



The data (texts and answers) are prepared in such a way that the task deliberately exhibits large ambiguity and variation, as well as coverage of long tail phenomena by including a substantial amount of low-frequent, local events and entities.

- IMPORTANT DATES:**
- ⇒ August 14, 2017 Trial data ready
 - ⇒ December 1, 2017 Test data ready
 - ⇒ January 10-31, 2018 Evaluation
 - ⇒ February 28, 2018 Paper submission due
 - ⇒ April 30, 2018 Camera ready version
 - ⇒ Summer 2018 SemEval workshop

Interested? → www.understandinglanguagebymachines.org/longtailqa_se18.