## **Optimizing Nugget Annotations with Active Learning**

Gaurav Baruah, Haotian Zhang, Rakesh Guttikonda, Jimmy Lin, Mark D. Smucker, Olga Vechtomova **University of Waterloo** 

**PROBLEM:** Nugget-based evaluations require exhaustive *matching* of nuggets against system responses.



**SOLUTION:** Use Continuous Active Learning for Technology-Assisted Review of candidate matches.



Review L match candidates; re-train respective nugget classifiers after each review decision

Review **M** match candidates for the sentence; re-train respective nugget classifiers on review

## **EVALUATION VIA SIMULATION:** Compute *effort* over simulated assessment interfaces for strategies.

**Nugget:** the train crashed at the buffer stop

Sentence: A packed train has slammed into a barrier at a Buenos Aires station, killing 49 people and injuring hundreds of morning commuters.

## **Next Match Candidate**

- Present assessor with Most Likely Candidate Match pair.
- Assessor reads sentence (*effort*  $\lambda_{read}$ ).
- Assessor renders judgement (*effort*  $\lambda_{match}$ ) including 3. annotations for a positive match.
- 4. Assessor moves on to the next match candidate.







$$effort = \lambda_{read} + \lambda_{match} + (M - 1) \lambda_{match'}$$





**ACKNOWLEDGMENT:** Support provided in part by a CIKM'16/SIGIR Travel Grant.