

Text Analysis Conference TAC 2018



Sponsored by: **NIST**



Hoa Trang Dang

U.S. National Institute of Standards and Technology

Outline

- Intro to Text Analysis Conference (TAC)
- Intro to (all new) TAC 2018 tracks
 - Drug-Drug Interaction Extraction from Drug Labels (DDI)
 - Systematic Review Information Extraction (SRIE)
 - Streaming Multimedia Knowledge Base Population (SM-KBP)

TAC Goals

- To promote research in NLP based on large common test collections
- To improve evaluation methodologies and measures for NLP
- To build test collections that evolve to meet the evaluation needs of state-of-the-art NLP systems
- To increase communication among industry, academia, and government by creating an open forum for the exchange of research ideas
- To speed transfer of technology from research labs into commercial products

Features of TAC

- Component evaluations situated within context of end-user tasks (e.g., summarization, knowledge base population)
 - opportunity to test components in end-user tasks
- Test common techniques across tracks
- “Small” number of tracks
 - critical mass of participants per track
 - sufficient resources per track (data, annotation/assessing, technical support)
- Leverage shared resources across tracks (organizational infrastructure, data, annotation/assessing, tools)

Workshop

- “Working workshop” – audience participation encouraged
- Presenting work in progress
- Extensive time scheduled for discussion/planning, esp. SM-KBP
- Targeted audience is participants in the shared tasks and evaluations, objective is to improve system performance
 - Improve evaluation specifications and infrastructure
 - Discuss and investigate intriguing/unexpected evaluation results
 - Learn from other teams

TAC 2018 Track Participants – THANK YOU

- DDI Track Coordinators
 - NIH/NLM (Dina Demner-Fushman)
- SRIE Track Coordinators
 - NIH/NIEHS (Mary Wolfe, Charles Schmitt)
- SM-KBP Organizers
 - NIST (Hoa Dang, Shahzad Rajput, George Awad, Asad Butt, Oleg Aulov, Ian Soboroff)
 - LDC (Stephanie Strassel, Jennifer Tracey, Jeremy Getman)
 - MITRE (Jason Duncan)
 - DARPA (Boyan Onyshkevych, Joseph Olive, Caitlin Christianson)
- EDL Track Coordinators (eval postponed to TAC 2019)
 - Heng Ji (RPI) and Avi Sil (IBM)
- 25 teams participating

Drug-Drug Interaction Extraction from Drug Labels

- Task 1: Extract Mentions of Interacting Drugs/Substances, interaction triggers and specific interactions at sentence level. This is similar to many NLP named entity recognition (NER) evaluations.
- Task 2: Identify interactions at sentence level, including: the interacting drugs, the specific interaction types (pharmacokinetic, pharmacodynamic or unspecified), and the outcomes of pharmacokinetic and pharmacodynamic interactions. This is similar to many NLP relation identification evaluations.
- Task 3: Normalization task. Normalize the interacting substance, drug classes, consequences of the interaction (if it is a medical condition), and pharmacokinetic effects.
- Task 4: Generate a global list of distinct interactions for the label in normalized form.

Systematic Review Information Extraction

- Goal is to develop and evaluate IE approaches that can assist in systematic reviews from published scientific literature.
- Focus on IE of study design factors found in the Methods and Materials section of published studies of experimental animals exposed to environmental chemicals.
- Task 1: Identify experimental design factors for the categories of exposure, animal group, dose group, and endpoint, and apply the appropriate annotation tag.
- Task 2: Identify relations between experimental design factors from Task 1, assign the factors to groups, and apply the appropriate annotation tag.

Streaming Multimedia KBP

- DARPA AIDA (Active Interpretation of Disparate Alternatives) program evaluation
- Goal is to develop a multi-hypothesis semantic engine that generates explicit alternative interpretations of events, situations, and trends from a variety of unstructured sources, for use in noisy, conflicting, and potentially deceptive information environments.
- Task 1: TA1 extracts knowledge elements from a stream of heterogeneous documents containing multilingual multimedia sources
- Task 2: TA2 aggregates the knowledge elements (from TA1) from multiple documents without access to the raw documents themselves (maintaining multiple interpretations and confidence values for KEs extracted or inferred from the documents)
- Task 3: TA3 develops semantically coherent hypotheses, each of which represents an interpretation of the document stream.

TAC 2019 Preview

- Systematic Review Information Extraction (SRIE)
- Drug-Drug Interaction from Drug Labels (DDI)
- Streaming Multimedia KBP (SM-KBP)
- Entity Discovery and Linking (EDL)