

# Text Analysis Conference TAC 2016



Sponsored by: **NIST**



Hoa Trang Dang

*National Institute of Standards and Technology*

# TAC 2017++ Session

- TAC 2017:
  - Adverse Drug Reaction Extraction from Drug Labels (Dina Demner Fushman, NIH/NLM/LHC)
  - KBP:
    - Cold Start++ KB Construction task
    - Component tasks: EDL; SF; EAL; EN Detection and coreference; Belief and Sentiment
    - *(Tentative)* Event Sequencing Pilot
- Panel: “What Next, After 2016?”
  - Generate ideas, plans for tasks for 2018 and beyond
- Broad Call for track proposals for TAC 2018
  - *All tracks must submit a written track proposal*

# KBP 2017

- Composite Cold Start++ KB Construction task (Required of DEFT teams)
  - Systems construct KB from raw text. KB contains:
    - Entities
    - Relations (Slots)
    - Events
    - Some aspects of Belief and Sentiment
  - KB populated from English, Chinese, and Spanish (30K/30K/30K docs)
- Component KBP tasks (as in 2016)
  - EDL
  - Slot Filling
  - Event Argument Extraction and (within-doc) Linking
  - Event Nugget Detection and (within-doc) Coref; Event Sequencing (tentative)
  - Belief and Sentiment

# Cold Start ++

- Minimize changes to existing KBP tasks and evaluation paradigms – change just enough to “bring it all together” into a single KB
  - Use existing evaluation/assessment tools as much as possible
  - Use existing input/output format as much as possible for each component
- Approach: Start with Cold Start 2016 KB, extend as needed to include Events and Belief/Sentiment.
- Each team submits a full KB, and we extract each **component** and evaluate as in 2016
- Additional **composite** score for KB: Extend Cold Start queries (currently limited to slot filling queries) to include event argument queries and sentiment queries

# Component evaluations for 2017

- EDL evaluation via ERE annotations + cross-doc entity coref (same as 2016)
- SF evaluation via assessment of selected queries (same as 2016)
- Event Nugget evaluation:
  - within-doc detection and coreference evaluation via ERE annotations (same as 2016)
  - subsequencing evaluation via ERE + annotation of after-links and parent/child links
- Event Argument evaluation: within-doc Event ARG extraction and linking via ERE gold standard annotation (same as 2016)
- Best evaluation via BeSt annotation over ERE gold standard annotation

# KBP 2017 Evaluation Windows

- June 30 - July 28: Cold Start++ KB Construction
- July 14 – July 28: Slot Filling
- Late September (TBA): EDL, EAL, EN
- Early October (TBA): Event sequencing, BeSt

# KB Entities

- Same schema as in CS2016 KB
- PER, ORG, GPE, FAC, LOC
- All NAM, NOM mentions; optional PROnominal mentions
- Only specific, individual entities (no unnamed aggregates)
  - “3 people” treated as a string value if it appears as an event argument; KB doesn’t need to extract or attempt to link \*all\* mentions of these aggregates
- + Require node ID to match entity node in the reference KB if linkable

:m.050v43	type	PER		
:m.050v43	mention	“Bart Simpson”		Doc1:37-48
:m.050v43	nominal_mention	“brother”		Doc1:15-21
:m.050v43	canonical_mention	“Bart Simpson”		Doc1:37-48

# KB Relations (Slot Filling)

- Same schema as in CS2016 KB

:e4 per:siblings :e7 Doc2:283-288,Doc2:173-179 0.6

:e4 per:siblings :e7 Doc3:283-288,Doc3:184-190 0.4

- But, for each justification, require all justification spans to come from the same document
- Assess  $k \geq 2$  justifications for each relation (for KBs only, not for runs submitted to standalone SF task)
- Make MAP the primary metric

# Assess more than one justification per relation

- Allow and assess up to  $k \geq 2$  justifications per relation for KBs
  - (Allow only one justification per relation for SF runs)
  - Each justification can have up to 3 justification spans; all spans must come from the same document
    - Multi-doc text spans in provenance allow more inferred relations => Perhaps put provenance for inference into separate column
- Justification1 is different from Justification2 iff justification spans come from different documents
- Credit for a Correct relation is proportional to number of different documents returned in the set of Correct justifications

# MAP and multi-hop confidence values

- Add Mean Average Precision (MAP) as a primary metric to consider confidence values in KB relation justifications
- To compute MAP, rank all responses (single-hop and multi-hop) by confidence value
  - Hop0 response: confidence is same as confidence associated with that justification
  - Hop1 response: confidence is product of confidence of each single-hop response along this path (from query to hop1)
  - Errors in hop1 get penalized less than errors in hop0
  - MAP could be a way to evaluate performance on hop0 and hop1 in a unified way that doesn't overly penalize hop1 errors.

# Event Nugget

- EN 2016 Nugget:

- doc1 E1 429,434 death lifedie actual
- doc1 E8 1420,1424 late lifedie actual

- EN 2016 Coreference

- HOPPERdoc1\_1 E1,E8

- EN attaches event.type.subtype to event nugget, but in KB we'll attach it to the event hopper

- Unlike ERE, subtypes of Contact and Transaction mentions must match in order to be coreferenced in KB

- CS2017:

- :Event1 type LIFE.DIE
- :Event1 mention.actual "death" doc1:429-433 # note difference in end offset
- :Event1 mention.actual "late" doc1:1420-1423
- :Event2 mention.other "die" doc1:34-36

- Don't evaluate cross-doc event nugget coreference in component evaluation

# Event Arguments in CS++

- EAL 2016 argument file: Each line is an assertion of an event argument (including event type, role, justifications, realis, confidence), with a unique ID
  - TFRFdoc1\_9 doc1 Life.Die Victim Zhou Enlai 1491-1500 1393-1500 1491-1494 NIL Actual 0.9
- EAL 2016 linking file:
  - HOPPERdoc1\_1 TFRFdoc1\_9,TFRFdoc1\_66
  - HOPPERdoc1\_2 TFRFdoc1\_22,TFRFdoc1,89
- EAL 2016 corpusLinking file
  - HOPPER\_1 HOPPERdoc1\_1,HOPPERdoc2\_3
- **CS++ 2017: Reify event hopper and reformat EAL justifications to look like CS SF justifications**

# BeSt

- What targets in the KB can be BeSt targets?
  - Entity targets
    - sentiment from entity to entity fits naturally into KB (sentiment slot filling in KBP 2013-2014)
  - *Don't allow Relations as targets in KB*
    - very few ERE relations are targets for sentiment
    - most ERE relations are targets for belief, but they're almost all CB
    - Relations/slots in Cold Start KB are supposed to be ACTUAL, highly probable
  - *Don't allow Events as targets in KB*
    - *Automatic event processing may not be mature enough to provide usable input to BeSt*

# Sentiment from entity towards entity

- Treat like regular relation (slot), but allow only one justification span per provenance,
- Justification is a mention of the target entity. Source must have a mention in the same document
- Return all justifications for each sentiment relation
- We evaluate justifications and sentiment relations in sample of docs

:e4	per:likes	:e7	Doc3:173-179	0.8
:e4	per:likes	:e7	Doc4:183-189	0.9
:e4	per:dislikes	:e7	Doc5:273-279	0.4
:e4	per:dislikes	:e8	Doc6:173-179	0.6
:e4	per:dislikes	:e8	Doc7:184-190	0.4

# COMPOSITE KB eval

- ***Evaluate entire KB by assessment of entity-focused queries***
- Ideally, sample queries to balance slot types, sentiment polarity, event types+roles (large number of sparse categories)
  - Queries may need to exclude some event types or event roles completely
- Score for interesting/complex queries is likely to be vanishingly small
  - Possibly use some derived queries (sampled from each submitted KB)

# Event Subsequence Linking Tasks for English in 2017 (tentative)

- Goal: Extract Subsequence of events
  - Input: Event nugget annotated files
  - Outputs: (1) After links; (2) Parent-Child links
- Corpus: Newswire and Discussion Forum in English
- Training data and Annotation Guidelines will be available for interested participants
  - Annotation tool: Modified Brat tool
- Scorer, submission validation scripts and submission format will be created by CMU