TAC KBP Event Detection Evaluation

Version 8.1, September 30, 2014

Carnegie Mellon University
Contact: Teruko Mitamura teruko@cs.cmu.edu

Overview

The Event Detection task at NIST TAC KBP aims to identify the explicit mentioning of Events in text. Every instance of a mention of the relevant Event types must be identified. The eventual benefit of the Event Detection Task is to detect full and partial coreference for the future task.

This task is required to detect the Event Types and Subtypes taken from the ERE guidelines version 3.1. Also, the task is to identify three REALIS {ACTUAL, GENERIC, OTHER}, which are described in the **Event Detection Annotation Guidelines**. The data sources are provided by LDC. About 150 annotated corpora will be provided prior to the evaluation as a training set. For the formal evaluation, about 200 corpora will be given to the participants. We plan to include newswire articles and discussion forums.

The Event Detection Task requires participants to identify all relevant Events Mention instances within each sentence. If the same Event is mentioned in several places in the document, the participants will list them all.

Task

For the Event Detection Task, participating systems will extract the following items:

1. Event Type and Subtypes (listed below)
2. REALIS Value (one of: ACTUAL, GENERIC, OTHER)
3. Event Nugget Identification (offset and character length)
4. Confidence Scores for 2 and 3 (optional).

1. Event Type and Subtypes Detection

For purposes of this evaluation, an event must fall under one of the event types and subtypes below. For more details, see the **Event Detection Annotation Guidelines.**

TYPE	**SUBTYPE**	**TYPE**	**SUBTYPE**
Life	Be-Born	Transaction	Transfer-Ownership
Life	Marry	Transaction	Transfer-Money
Life	Divorce	Justice	Arrest-Jail
Life	Injure	Justice	Release-Parole
Life	Die	Justice	Trial-Hearing
Movement	Transport-Person	Justice	Charge-Indict
Business	Start-Org	Justice	Sue
Business	End-Org	Justice	Convict
Business	Declare-Bankruptcy	Justice	Sentence
Business	Merge-Org	Justice	Fine
Conflict	Demonstrate	Justice	Execute
Conflict	Attack	Justice	Extradite
Contact	Meet	Justice	Acquit
Contact	Communicate	Justice	Appeal
Personnel	Start-Position	Justice	Pardon
Personnel	End-Position		
Personnel	Nominate		
Personnel	Elect		

2. REALIS Identification

Event mentions will refer to **ACTUAL** (events that actually occurred); **GENERIC** (events that are not specific events with a (known or unknown) time and/or place); or **OTHER** (which includes failed events, future events, and conditional statements, and all other non-generic variations). For more detail explanation, see the **Event Detection Annotation Guidelines.**

3. Event Nugget Identification

A system will choose to identify Event Nuggets in the text. The definition of the Event Nugget generally follows the **Event Detection Annotation Guidelines**. Each Nugget is the actual string of words that indicate the mentioned event, and must correspond to the Event type and subtype above. When a sentence/block mentions more than one event type both must be mentioned, e.g., in the example sentence "he shot the soldier dead," both [conflict.ATTACK] and [life.DIE] are events. For more details, see the separate document **Event Detection Annotation Guidelines**.

System Output

For each nugget detected, the system must output one line in a text file:
If the system chooses not to provide the confidence scores, the last two fields are empty. Details of system output formats are described in the **Event Mention Detection Scoring** document.

System Outputs:

- system-ID: unique ID assigned to each system run
- doc-ID: unique ID assigned to each source document
- mention ID: ID of the event nugget
- token ID list: list of IDs for the token(s) of the current mention
- mention-string: character string of event mention (see **Event Detection Annotation Guidelines**)
- event-type: *type.subtype* from the hierarchy given above
- Realis-value: one of *ACTUAL, GENERIC, OTHER*
- Confidence scores of event type: score between 0 and 1 inclusive (optional)
- Confidence scores of Realis-value: score between 0 and 1 inclusive (optional)

Corpus

The corpus for this task will consist of 200 documents from two different types of documents: newswire and discussion forum documents. About half will be taken from each genre. The documents are in XML format and may be segmented into blocks.

Scoring

Systems will be scored using precision and recall over the gold standard.

The scorer reads the output of event mention detection systems and compares them to the gold standard.

Input:

Gold standard annotation for a text, in format (one line per mention)
System output annotation for the same text, in format (one line per mention)

Output:

System output annotation for the same text, with score appended to each line

The detail scoring definition is described in the separate document: **Event Mention Detection Scoring v15.**

Submissions and Schedule

Systems will have up to one week to process the evaluation documents. Submissions should be fully automatic and no changes should be made to the system once evaluation corpus has been downloaded. Up to **three** alternative system runs may be submitted per-team. Submitted runs should be ranked according to their expected overall score.

Sept 2, 2014	Release Event Mention Task definition
Sept 2, 2014	Release a draft of Annotation Guidelines
Sept 19, 2014	Release Final Annotation Guidelines
Sept 30, 2014	LDC releases of Pilot Annotated Training data (20 docs) LDC releases sample documents (20 docs)
Oct 21, 2014	Participants (optionally) submit pilot output (1~20 docs from sample documents) for pilot assessment. Pilot may be automatic, manual, or some combination. While participation is not required, participants are encouraged to submit output to better understand this new task (e.g. receive feedback on the assessment of event nugget, event types/subtypes and Realis).
Nov 4, 2014	The pilot assessment to be released to participants who submitted pilot output.
Nov 26, 2014	Release of Annotated Training data (100 docs)
Jan 16-22, 2015	Evaluation data released to performers
Jan 23-30, 2015	System output to NIST
Feb 13, 2015	Return the results to performers