# TAC KBP Event Mention Detection Evaluation

Version 7.4, August 5, 2014

Carnegie Mellon University

Contact: Teruko Mitamura [teruko@cs.cmu.edu](mailto:teruko@cs.cmu.edu)

**Overview**

The Event Mention Detection task at NIST TAC KBP 2014 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 Mention 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 Mention 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 Mention 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 Mention 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 Mention Identification (offset and character length)
4. Confidence Scores for 2 and 3 (optional).

Systems will emit sentences or blocks that contain a mention of an event in the ERE ontology of events. The output will be a text file with lines of 5 fields each: ‘Block ID’ is a text location identifier; ‘Event type/Subtype’ is a name from the ERE v1.3 event ontology (for example, Conflict.ATTACK); the REALIS value selected for the mention; and the event trigger (a text string) as the system’s evidence for the Event Mention detected. ‘Confidence Score’ is a real number in the interval [0,1] that indicates the system’s confidence, belief, probability, etc., in the identification of that sentence as mentioning that event; Details are provided below.

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 Mention 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 |  |  |  |

1. **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 Mention Annotation Guidelines.**

1. **Event Mention Identification**

A system will choose to identify Event Mentions in the text. The definition of the Event Mention generally follows the **Event Mention Annotation Guidelines**. Each Mention 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 Mention Annotation Guidelines**.

**System Output**

For each mention detected, the system must output one line in a text file:

If the system chooses not to provide the confidence scores, use a hyphen sign “-“ at the place of <score 1> (i.e. hyphen represent an empty field).

Each line of event mention is formatted as followed:

<system ID><TAB><doc ID><TAB><mention ID><TAB><token ID list><TAB> <mention><TAB><event-type><TAB><realis status><TAB><score1><TAB> <score2><TAB><score3>

**Explanation:**

<system ID> := the name of the system

<doc ID> := the ID of the input document

<mention ID> := the ID of the mention, which should uniquely identify the mention within the current document

<token ID list> := list of IDs for the token(s) of the current mention, in ascending order, separated by commas (,)

<mention> := the actual character string of the mention

<event-type> := the ACE hierarchy type

<realis status> := the REALIS label

<score1> := any score (confidence, etc.) the system wants to assign, if ignored, please put a hyphen (-) to indicate an empty field

<score2> := score assigned in the evaluation

<score3> := additional possible score assigned by human

<TAB> := tab character

**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**

The scorer reads the output of event mention detection systems and compares them to the gold standard. The scorer will create a one-to-many mapping from gold standard mentions to system mentions and provide mapping score for each mapping pair accordingly.

The scorer will then create partial scores. Event mention detection scores are calculated by choosing the best mapping for each gold standard mention. Mention type and realis status detection are calculated by assigning proportional score to each possible mapping.

Overall scores will be created by aggregating such partial scores.

**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:**

Gold standard annotation for the same text, with the system score for each gold standard mention appended to each line

A summarized performance report that contains score for each mention, each document, and the whole corpus

For detail documentation on scoring and examples, please refer to the attached scoring document (Event-Mention-Detection-scoring.docx).

**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.

|  |  |
| --- | --- |
| August 15 2014 | Release Event Mention Task definition |
| August 29, 2014 | Release a draft of Annotation Guidelines |
| Sept 4, 2014 | LDC/NIST releases sample documents (~20 docs) to participants in preparation for pilot assessment |
| Sept 19, 2014 | Release Final Annotation Guidelines |
| Sept 22, 2014 | Participants (optionally) submit pilot output (~20 docs) 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 mention, event types/subtypes and realis). |
| Sept 24, 2014 | Release of Pilot Annotated Training data (30 docs) |
| Oct 8 2014 | The full set of pilot assessment to be released to participants who submitted pilot output. |
| Nov 26, 2014 | Release of Annotated Training data (100 docs) |
| Jan 2-8, 2015 | Evaluation data released to performers |
| Jan 9-16, 2015 | System output to NIST |
| Jan 30, 2015 | Return the results to performers |