

Addendum for ALADDIN Performers TRECVID Multimedia Event Recounting (MER) Evaluation Plan

BACKGROUND

This addendum identifies the *additional requirements* for ALADDIN performers participating in the 2012 TRECVID MER evaluation. The requirements and assessments covered in this document are not available to non-ALADDIN MER participants.

SYSTEM TASK

The system task is identical to the one specified in the TRECVID Multimedia Event Recounting (MER) Evaluation Plan, except that the ALADDIN participants will be required to track and report a total of three times (instead of the two times in the TRECVID Eval Plan):

- (1) the time required for CDR generation,
- (2) the time required for Event Agent execution
(that is, evidence identification and extraction), and
- (3) the time required for MER output generation.

All other descriptions of the task remain the same.

MER-2012 EVENTS

There are no additional changes from those listed in the TRECVID MER evaluation plan.

DATA RESOURCES

There are no additional changes from those listed in the TRECVID MER evaluation plan.

DRY RUN

There are no necessary changes to the dry run to test the implementation and assessments specified here.

EVALUATION

These assessments will be performed for each ALADDIN system.

The first additional test will measure how well the MER systems capture the most critical pieces of evidence present in the clip.

Prior to the evaluation, and using only the information in the event kits, each judge will view each clip designated as a positive for the MER evaluation, and will identify the top few (up to five) pieces of evidence that indicate that the clip contains an instance of the event. The judges will use these (personal) lists of evidence when judging the MER outputs.

For assessment of the output of a MER system, each judge will indicate whether the top pieces of evidence (from his/her personal list) are present in the recounting (MER system output).

The second additional test will measure qualitative opinions using six Likert-type questions about each MER recounting. For each question, the responses will be on a five-point scale of Strongly Agree through Strongly Disagree.

1. The recounting accurately described the event as it occurred in the clip.
2. The recounting was concise.
3. The recounting was easily comprehended.
4. The recounting addressed the important evidence in the event kit.
5. The recounting included relevant evidence above and beyond what was specified in the event kit.
6. The confidence scores actually did correspond to the accuracy of the items in the recounting.

METRICS

Metrics for determining how well MER systems capture the most critical pieces of evidence:

For each recounting from each system, two metrics will be calculated for the critical pieces of evidence.

- Recall: The mean (across all judges) of the fraction of the pieces of evidence on the judge's personal list of critical pieces of evidence that the judge decided were present in the recounting (the "recalled evidence").
- Confidence: The mean confidence factor across all recalled pieces of evidence.

For each system, we will also report the mean across all recountings for these two metrics.

Metrics for measuring qualitative opinions:

The five-point Likert-type type responses from the judges will be as follows.

Strongly agree

Agree

Neutral

Disagree

Strongly Disagree

For purposes of scoring, we will convert these responses to numbers (+2 for *Strongly agree*, through -2 for *Strongly disagree*).

As the main metric, for each recounting from each system we will report the mean response across all judges for each of the six questions.

To assist in understanding the judges responses, we will also compute statistical measures of dispersion and location, and we will also provide a presentation of the results as categorical data..

SCHEDULE

There are no changes to the TRECVID schedule.

INPUTS/OUTPUTS

There are no changes to the system input/outputs. Aladdin performers will, however, be expected to implement the full range of elements and attributes from the DTD.