

NIST Rich Transcription Evaluation 2007 Evaluation

Registration Agreement Form

To participate in the 2007 NIST Rich Transcription Evaluation (RT-07) this registration form must be filled out in its entirety and returned to NIST by the registration deadline of 7-February-2007

FAX: Attention Jonathan Fiscus
+1 (301) 670-0939

or

Postal Mail: Jonathan Fiscus - NIST
100 Bureau Dr. Mail Stop 8940
Gaithersburg, MD 20899-0940 USA

The following are terms of participation in the RT-07 evaluation:

- On receiving the evaluation data (21-February-2007), the registered site will process each test set in its entirety in accordance with the RT-06S Evaluation Plan, and return the system output files to NIST by the end of the evaluation period (21-March-2007 5:00 pm EDT for STT and SPKR tasks).
- The evaluation source data and ancillary annotations are not to be examined until such time that all RT evaluations have been completed.
- The site agrees to be represented at the Rich Transcription Evaluation Workshop to be held in Baltimore, Maryland May 10-11, 2007, and to openly discuss the system(s) evaluated. Participants must write a system description document that will be included in the workshop materials and deliver a prepared system description presentation at that workshop.
- The evaluation data cannot be redistributed. Sites may keep the data and continue using the data for research purposes.
- The site has registered with the Augmented Multi-Party (AMI) Interaction Program for free access to the AMI corpus by completing the agreements found on '<http://corpus.amiproject.org/>'.

The following are terms of publishing the NIST Rich Transcription evaluation results:

- Your organization agrees that the sponsoring organizations (U.S. Government and their designates) will be permitted to publish your results without anonymity or further permission in the RT Proceedings, RT Website, and relevant publications that the sponsoring organizations choose.
- Your organization understands that, although the evaluation is open to all wishing to participate, results are only discussed within the group of registered participants and representatives of supporting government agencies.
- Your organization agrees to not publicly compare its results with the results of other participants, such as stating rank ordering or score difference in a particular evaluation task. Your organization is free to do what they wish with their own results, but may neither redistribute nor publish results from another site without that site's explicit permission. However, organizations are encouraged to reference the RT website and official RT evaluation publications in their publications.
- This agreement does not supersede any contractual agreement that a site may have with its sponsors or their affiliates regarding publication review.

NIST serves to coordinate these evaluations. These tests are designed for local implementation by each participant. As such, this evaluation is based on an honor system. Organizations that are found to have manually or artificially generated results for these evaluations may be barred from future participation in any RT-affiliated evaluations.

The reported results are not to be construed, or represented, as endorsement of any participant's system, or as official findings on the part of NIST, the U.S. Government.

Initials: _____

NIST Rich Transcription Evaluation 2007 Evaluation

(Continued)

Site Affiliation: _____

Evaluation Tasks: (check boxes of planned participation)

	Conference Room Data	Lecture Room Data	Coffee Break Data	Broadcast News
Speech-To-Text				
Speaker Diarization				
Speaker Attributed STT (non-binding intent to participate)				
Text Recognition in Video				

Contact Information

Name: _____ Address for evaluation data: _____

Phone: _____

Signature: _____

Date: _____

By signing and returning this form to NIST you are registering your site as a participant in the NIST RT-06S evaluation, and acknowledging that you have read and will abide by the protocols as described here and in the official evaluation plan.

NIST Rich Transcription Evaluation

2007 Evaluation

Data Availability Addendum

This form identifies the corpora available to system developers for the 2007 NIST Rich Transcription Evaluation (RT-07). After signing and returning this form along with the RT-07 evaluation agreement, the data distribution agencies will be notified of the participant's eligibility to receive the data. While some of the resources below are publicly available through the Linguistic Data Consortium (LDC) and/or the Evaluations and Language resources Distribution Agency (ELDA), others below are only distributed through special arrangements such as this evaluation. Participants agree to use all corpora in accord with the licenses accompanying each resource.

Publicly available meeting resources:

- ICSI Meeting Speech: LDC catalog number LDC2004S02
- ICSI Meeting Speech: LDC catalog number LDC2004T04
- ISL Meeting Speech Part1: LDC catalog number LDC2004S05
- ISL Meeting Transcripts Part 1: LDC catalog number LDC2004T10
- NIST Meeting Pilot Corpus Speech: LDC catalog number LDC2004S09
- NIST Meeting Pilot Corpus Transcripts and Metadata: LDC catalog number LD2004T13
- RT-04S Dev-Eval Meeting Room Data (speech+transcripts) LDC2005S09
- Rich Transcription 2005 Spring Evaluation Data
- Rich Transcription 2006 Spring Evaluation Data
- Effective, Affordable, Reusable, Speech-To-Text (EARS) RT-04 Broadcast News training corpus distributed to non-EARS partners as a resource for developing RT-06S systems.
 - Topic Detection and Tracking Phase 4 (TDT4) Multilingual Broadcast News Speech Corpus: LDC2005S11
 - TDT4 Multilingual Text and Annotations: LDC2005T16
- Effective, Affordable, Reusable, Speech-To-Text (EARS) RT-04 Conversational Telephone Speech training corpus distributed to non-EARS partners as a resource for developing RT-06S systems:
 - Fisher English Training Speech Part 1 Speech: LDC catalog number LDC2004S13 (5850 two sided telephone conversations)
 - Fisher English Training Speech Part 1, Transcripts: LDC catalog number LDC2004T19 (5850 transcribed two sided telephone conversations)
 - Fisher English Training Speech Part 2: LDC catalog number LDC2005S13
 - Fisher English Training Speech Part 2, Transcripts: LDC catalog number LDC2005T19

Other corpora offered to the RT-07 evaluation participants:

The corpora listed in this section have been produced by several non-affiliated programs. A data sharing agreement has been reached whereby sites not affiliated with each corpus' producer are granted a non-transferable evaluation license to the data. Sites are allowed to retain and use the data for research purposes.

- Computers in the Human Interaction Loop (CHIL) development test set: a five meeting data set collected by the CHIL Consortium and distributed to non-CHIL partners as a resource for developing RT systems.
- CHIL Development 2006 Corpus - Includes the RT-05S Lecture Room test set and interactive seminars.
- Augmented Multiparty Interaction (AMI) development test set: a twelve meeting data set collected by the AMI project and distributed to non-AMI partners as a resource for developing RT systems.

By signing this form, the site acknowledges they have returned to NIST a completed RT-06S Evaluation Agreement form and agrees to the terms of use for the aforementioned non-publicly available resources.

Site Affiliation: _____

Signature: _____

Date: _____