

# **2008 NIST Open Machine Translation Evaluation (MT08)**

## **Human Assessments – Form**

Site participation in the human assessment portion of MT08 requires MT08 Form 4, Human Assessments Participation Agreement, to be completed and signed by the site's Principal Investigator. The form must be returned by fax no later than **January 10 2008** to:

NIST  
Attn: NIST Open MT Team  
Fax +1 (301) 670-0939

### **Introduction**

NIST will coordinate the MT08 human assessments of translation adequacy and preference for a subset of the Current Test Set for the Arabic-to-English, Chinese-to-English, and Urdu-to-English tracks. Per language track, each participant can choose to have their primary system for the Constrained and/or Unconstrained training condition(s) included in the assessments.

Site participation in the human assessments is voluntary, and is contingent on providing volunteer judges to assist in making the assessments. All decisions made by the volunteer judges represent their own opinion, not that of their affiliation. All judges may only be recruited on a volunteer basis.

Each site that wishes to have their system translations included in the human assessment process must meet the requirements for participation specified in the Human Assessments Participation Agreement (MT08 Form 4).

For each site that agrees to participate, a subset of the site's system translations from the Current Test Set in the selected training and language track(s) will be added to the human assessment data pool.

If a site meets the requirements for participation, the human assessment scores for the site will be included in reports of results and the site will receive the raw scores with the participating sites identified. If a site agrees to participate but subsequently fails to meet the requirements for participation, the site's system's human assessment scores will be extracted and discarded from all reports and distributions.

To better gauge the time and effort necessary to meet the requirements fully, it is strongly recommended to explore the required tutorial ([http://www.nist.gov/speech/tests/mt/ha\\_overview/](http://www.nist.gov/speech/tests/mt/ha_overview/)) before agreeing to participate.

# 2008 NIST Open Machine Translation Evaluation (MT08)

## Human Assessments – Form

### NIST OPEN MT 2008 Human Assessments Participation Agreement – MT08 Form 4

I affirm that (print site affiliation): \_\_\_\_\_

chooses to include its primary\* MT system for (check all that apply):

Arabic to English:	<input type="checkbox"/> Constrained	<input type="checkbox"/> Unconstrained
Chinese to English:	<input type="checkbox"/> Constrained	<input type="checkbox"/> Unconstrained
Urdu to English:	<input type="checkbox"/> Constrained	

in the MT08 human assessments. A portion of my site's system translations from the Current Test Set for each selected language/training track combination will be included in the data set to be assessed by a pool of judges from different participating sites and non-participants.

\* By default, translations from the primary system will be used, but there will be a two day period after preliminary results are released during which a site may identify an alternative selection for inclusion in the human assessment study.

To participate, I agree to fulfill these **minimum requirements** for participation:

For **each** language/training track combination selected above, provide at least one native or near-native English speaking judge to complete, no later than **March 16 2008**, all of:

- Review of the web application overview ([http://www.nist.gov/speech/tests/mt/ha\\_overview/](http://www.nist.gov/speech/tests/mt/ha_overview/)), and
- Example set: approximately 20 adequacy judgments, approximately 20 preference judgments, and
- Evaluation set: approximately 200 adequacy judgments, approximately 200 preference judgments.

NIST reserves the right to reject a judge if:

- The example set of assessments by a judge indicate insufficient understanding of the assessment guidelines outlined in the tutorial, or
- The actual judgments indicate insufficient time to complete, or insufficient understanding of, the assessment guidelines.

Sites will be given the opportunity to replace judges or withdraw from the human assessments process.

My site will recruit judges on a strictly voluntary basis. Judges will provide their own opinion, not that of their affiliation. Judges will not be compensated.

NIST will neither request nor record any information personally identifying individual judges. My site will provide a point-of-contact e-mail address to which unique log-in IDs for the number of judges to be provided will be sent, and will forward one of those IDs to each judge. The log-in IDs will be linked to the site for the duration of the study (to track compliance with participation requirements), but not to individual judges. My site will not keep any records linking assessments to individual judges provided.

I understand that, if my site meets all requirements for participation, the human assessment scores for my site's system output will be included in reports of results and I will receive a distribution of the raw scores with the corresponding participants identified. If my site fails to meet the requirements for participation, the human assessment scores for my site's system translations will be extracted and discarded from reports of results and data distributions.

Site Principal Investigator name (print): \_\_\_\_\_

Site Principal Investigator signature: \_\_\_\_\_

Date: \_\_\_\_\_