

Workshop on Creating Speech and Language Data with Amazons Mechanical Turk 2010

Held at NAACL HLT 2010

**Los Angeles, California, USA
6 June 2010**

ISBN: 978-1-61738-676-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2010) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2010)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Creating Speech and Language Data With Amazon’s Mechanical Turk</i>	
Chris Callison-Burch and Mark Dredze	1
<i>Corpus Creation for New Genres: A Crowdsourced Approach to PP Attachment</i>	
Mukund Jha, Jacob Andreas, Kapil Thadani, Sara Rosenthal and Kathleen McKeown	13
<i>Clustering dictionary definitions using Amazon Mechanical Turk</i>	
Gabriel Parent and Maxine Eskenazi	21
<i>Semi-supervised Word Alignment with Mechanical Turk</i>	
Qin Gao and Stephan Vogel	30
<i>Rating Computer-Generated Questions with Mechanical Turk</i>	
Michael Heilman and Noah A. Smith	35
<i>Crowdsourced Accessibility: Elicitation of Wikipedia Articles</i>	
Scott Novotney and Chris Callison-Burch	41
<i>Document Image Collection Using Amazon’s Mechanical Turk</i>	
Audrey Le, Jerome Ajot, Mark Przybocki and Stephanie Strassel	45
<i>Using Amazon Mechanical Turk for Transcription of Non-Native Speech</i>	
Keelan Evanini, Derrick Higgins and Klaus Zechner	53
<i>Exploring Normalization Techniques for Human Judgments of Machine Translation Adequacy Collected Using Amazon Mechanical Turk</i>	
Michael Denkowski and Alon Lavie	57
<i>Can Crowds Build parallel corpora for Machine Translation Systems?</i>	
Vamshi Ambati and Stephan Vogel	62
<i>Turker-Assisted Paraphrasing for English-Arabic Machine Translation</i>	
Michael Denkowski, Hassan Al-Haj and Alon Lavie	66
<i>Annotating Large Email Datasets for Named Entity Recognition with Mechanical Turk</i>	
Nolan Lawson, Kevin Eustice, Mike Perkowitz and Meliha Yetisgen-Yildiz	71
<i>Annotating Named Entities in Twitter Data with Crowdsourcing</i>	
Tim Finin, William Murnane, Anand Karandikar, Nicholas Keller, Justin Martineau and Mark Dredze	80
<i>MTurk Crowdsourcing: A Viable Method for Rapid Discovery of Arabic Nicknames?</i>	
Chiara Higgins, Elizabeth McGrath and Laila Moretto	89
<i>An Enriched MT Grammar for Under \$100</i>	
Omar F. Zaidan and Juri Ganitkevitch	93

<i>Using the Amazon Mechanical Turk to Transcribe and Annotate Meeting Speech for Extractive Summarization</i>	
Matthew Marge, Satanjeev Banerjee and Alexander Rudnicky	99
<i>Using Mechanical Turk to Annotate Lexicons for Less Commonly Used Languages</i>	
Ann Irvine and Alexandre Klementiev	108
<i>Opinion Mining of Spanish Customer Comments with Non-Expert Annotations on Mechanical Turk</i>	
Bart Mellebeek, Francesc Benavent, Jens Grivolla, Joan Codina, Marta R. Costa-Jussà and Rafael Banchs	114
<i>Crowdsourcing and language studies: the new generation of linguistic data</i>	
Robert Munro, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen and Harry Tily	122
<i>Not-So-Latent Dirichlet Allocation: Collapsed Gibbs Sampling Using Human Judgments</i>	
Jonathan Chang	131
<i>Collecting Image Annotations Using Amazon's Mechanical Turk</i>	
Cyrus Rashtchian, Peter Young, Micah Hodosh and Julia Hockenmaier	139
<i>Non-Expert Evaluation of Summarization Systems is Risky</i>	
Dan Gillick and Yang Liu	148
<i>Shedding (a Thousand Points of) Light on Biased Language</i>	
Tae Yano, Philip Resnik and Noah A. Smith	152
<i>Evaluation of Commonsense Knowledge with Mechanical Turk</i>	
Jonathan Gordon, Benjamin Van Durme and Lenhart Schubert	159
<i>Cheap Facts and Counter-Facts</i>	
Rui Wang and Chris Callison-Burch	163
<i>The Wisdom of the Crowds Ear: Speech Accent Rating and Annotation with Amazon Mechanical Turk</i>	
Stephen Kunath and Steven Weinberger	168
<i>Crowdsourcing Document Relevance Assessment with Mechanical Turk</i>	
Catherine Grady and Matthew Lease	172
<i>Preliminary Experiments with Amazon's Mechanical Turk for Annotating Medical Named Entities</i>	
Meliha Yetisgen-Yildiz, Imre Solti, Fei Xia and Scott Halgrim	180
<i>Tools for Collecting Speech Corpora via Mechanical-Turk</i>	
Ian Lane, Matthias Eck, Kay Rottmann and Alex Waibel	184
<i>Measuring Transitivity Using Untrained Annotators</i>	
Nitin Madnani, Jordan Boyd-Graber and Philip Resnik	188
<i>Amazon Mechanical Turk for Subjectivity Word Sense Disambiguation</i>	
Cem Akkaya, Alexander Conrad, Janyce Wiebe and Rada Mihalcea	195

<i>Non-Expert Correction of Automatically Generated Relation Annotations</i>	
Matthew R. Gormley, Adam Gerber, Mary Harper and Mark Dredze	204
<i>Using Mechanical Turk to Build Machine Translation Evaluation Sets</i>	
Michael Bloodgood and Chris Callison-Burch	208
<i>Creating a Bi-lingual Entailment Corpus through Translations with Mechanical Turk: \$100 for a 10-day Rush</i>	
Matteo Negri and Yashar Mehdad	212
<i>Error Driven Paraphrase Annotation using Mechanical Turk</i>	
Olivia Buzek, Philip Resnik and Ben Bederson	217