

2nd Workshop on Human Evaluation of NLP Systems (HumEval 2022)

Dublin, Ireland
27 May 2022

ISBN: 978-1-7138-6739-5

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© (2022) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

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: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Vacillating Human Correlation of SacreBLEU in Unprotected Languages</i> Ahrii Kim and Jinhyeon Kim	1
<i>A Methodology for the Comparison of Human Judgments With Metrics for Coreference Resolution</i> Mariya Borovikova, Loïc Grobol, Anaïs Lefeuvre Halftermeyer and Sylvie Billot	16
<i>Perceptual Quality Dimensions of Machine-Generated Text with a Focus on Machine Translation</i> Vivien Macketanz, Babak Naderi, Steven Schmidt and Sebastian Möller	24
<i>Human evaluation of web-crawled parallel corpora for machine translation</i> Gema Ramírez-Sánchez, Marta Bañón, Jaume Zaragoza-Bernabeu and Sergio Ortiz Rojas	32
<i>Beyond calories: evaluating how tailored communication reduces emotional load in diet-coaching</i> Simone Balloccu and Ehud Reiter	42
<i>The Human Evaluation Datasheet: A Template for Recording Details of Human Evaluation Experiments in NLP</i> Anastasia Shimorina and Anya Belz	54
<i>Toward More Effective Human Evaluation for Machine Translation</i> Belén C Saldías Fuentes, George Foster, Markus Freitag and Qijun Tan	76
<i>A Study on Manual and Automatic Evaluation for Text Style Transfer: The Case of Detoxification</i> Varvara Logacheva, Daryna Dementieva, Irina Krotova, Alena Fenogenova, Irina Nikishina, Tatiana Shavrina and Alexander Panchenko	90
<i>Human Judgement as a Compass to Navigate Automatic Metrics for Formality Transfer</i> Huiyuan Lai, Jiali Mao, Antonio Toral and Malvina Nissim	102
<i>Towards Human Evaluation of Mutual Understanding in Human-Computer Spontaneous Conversation: An Empirical Study of Word Sense Disambiguation for Naturalistic Social Dialogs in American English</i> Alex Lutu	116