Tenth Workshop on Noisy User-generated Text (W-NUT 2025)

Albuquerque, New Mexico, USA
3 May 2025

ISBN: 979-8-3313-1932-8

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© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

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

Towards a Social Media-based Disease Surveillance System for Early Detection of Influenza-like Illnesses: A Twitter Case Study in Wales
Mark Drakesmith, Dimosthenis Antypas, Clare Brown, Jose Camacho-Collados and Jiao Song . 1

Sentiment Analysis on Video Transcripts: Comparing the Value of Textual and Multimodal Annotations
Quanqi Du, Loic De Langhe, Els Lefever and Veronique Hoste 10

Restoring Missing Spaces in Scraped Hebrew Social Media
Avi Shmidman and Shaltiel Shmidman 16

Identifying and analyzing 'noisy' spelling errors in a second language corpus
Alan Juffs and Ben Naismith 26

Automatic normalization of noisy technical reports with an LLM: What effects on a downstream task?
Mariame Maarouf and Ludovic Tanguy 38

We're Calling an Intervention: Exploring Fundamental Hurdles in Adapting Language Models to Nonstandard Text
Aarohi Srivastava and David Chiang 45

On-Device LLMs for Home Assistant: Dual Role in Intent Detection and Response Generation
Rune Birkmose, Nathan Mørkeberg Reece, Esben Hofstedt Norvin, Johannes Bjerva and Mike Zhang 57

Applying Transformer Architectures to Detect Cynical Comments in Spanish Social Media
Samuel Gonzalez-Lopez, Steven Bethard, Rogelio Platt-Molina and Francisca Orozco 68

Prompt Guided Diffusion for Controllable Text Generation
Mohaddeseh Mirbeygi and Hamid Beigy 78

FaBERT: Pre-training BERT on Persian Blogs
Mostafa Masumi, Seyed Soroush Majd, Mehrnoush Shamsfard and Hamid Beigy 85

Automatically Generating Chinese Homophone Words to Probe Machine Translation Estimation Systems
Shenbin Qian, Constantin Orasan, Diptesh Kanojia and Félix Do Carmo 97

Multi-BERT: Leveraging Adapters for Low-Resource Multi-Domain Adaptation
Parham Abed Azad and Hamid Beigy 108

Enhancing NER Performance in Low-Resource Pakistani Languages using Cross-Lingual Data Augmentation
Toqeer Ehsan and Thamar Solorio 117

Wikipedia is Not a Dictionary, Delete! Text Classification as a Proxy for Analysing Wiki Deletion Discussions
Hsuvas Borkakoty and Luis Espinosa-Anke 133

From Conversational Speech to Readable Text: Post-Processing Noisy Transcripts in a Low-Resource Setting
Arturs Znotins and Normunds Gruzitis 143

Text Normalization for Japanese Sentiment Analysis
Risa Kondo, Ayu Teramen, Reon Kajikawa, Koki Horiguchi, Tomoyuki Kajiwara, Takashi Ninomiya, Hideaki Hayashi, Yuta Nakashima and Hajime Nagahara 149