

WikiNLP 2025

**WikiNLP: Advancing Natural Language Processing for
Wikipedia**

Proceedings of the Workshop

August 1, 2025

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-284-8

Program Committee

Program Chairs

Akhil Arora, Aarhus University
Isaac Johnson, Wikimedia
Lucie-Aimée Kaffee, Hugging Face
Tzu-Sheng Kuo, Carnegie Mellon University
Tiziano Piccardi, Stanford University
Indira Sen, Universität Mannheim

Reviewers

Neelima Agarwal, Pablo Aragón, Hiba Arnaout, Wenceslao Arroyo-Machado

Hsuvas Borkakoty, Hannah Brueckner

Kaylea Champion

Djellel Difallah

Patrick Gildersleve, Dipanwita Guhathakurta, Lavanya Gupta

Steve Jankowski, Srihari Jayakumar, Deeptanshu Jha

Gaurav Kumar

Isabelle Langrock

Finn Årup Nielsen

Tiziano Piccardi, Karthik Sriranga Puthraya

Marija Sakota, Nicole Schwitter, Indira Sen, Andreas Spitz

Nathan TeBlunthuis, Carla Toro, Harold Triedman, Mykola Trokhymovych

Thejas Venkatesh

Jheng-Hong Yang, Jisung Yoon

Dale Zhou, Kai Zhu

Table of Contents

<i>Wikivecs: A Fully Reproducible Vectorization of Multilingual Wikipedia</i> Brandon Duderstadt	1
<i>WETBench: A Benchmark for Detecting Task-Specific Machine-Generated Text on Wikipedia</i> Gerrit Quaremba, Elizabeth Black, Denny Vrandečić and Elena Simperl	10
<i>Proper Noun Diacritization for Arabic Wikipedia: A Benchmark Dataset</i> Rawan Bondok, Mayar Nassar, Salam Khalifa, Kurt Micallef and Nizar Habash	31