

WikiNLP: Advancing Natural Language Processing for Wikipedia

Vienna, Austria
1 August 2025

ISBN: 979-8-3313-2421-6

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. (2025)

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>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 Vrandecic 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