4th Workshop on Arabic Corpus Linguistics (WACL-4)

Abu Dhabi, United Arab Emirates
20 January 2025

ISBN: 979-8-3313-1375-3

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© (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

ArabicSense: A Benchmark for Evaluating Commonsense Reasoning in Arabic with Large Language Models
Salima Lamsiyah, Kamyar Zeinalipour, Samir El Amrany, Matthias Brust, Marco Maggini, Pascal Bouvry and Christoph Schommer 1

Lahjawi: Arabic Cross-Dialect Translator
Mohamed Motasim Hamed, Muhammad Hreden, Khalil Hennara, Zeina Aldallal, Sara Chrouf and Safwan AlModhayan 12

Lost in Variation: An Unsupervised Methodology for Mining Lexico-syntactic Patterns in Middle Arabic Texts
Julien JB Bezançon, Rimane Karam and Gaël Lejeune 25

SADSLyC: A Corpus for Saudi Arabian Multi-dialect Identification through Song Lyrics
Salwa Saad Alahmari 38

Enhancing Dialectal Arabic Intent Detection through Cross-Dialect Multilingual Input Augmentation
Shehenaz Hossain, Fouad Shammary, Bahaulddin Shammary and Haithem Afli 44

Dial2MSA-Verified: A Multi-Dialect Arabic Social Media Dataset for Neural Machine Translation to Modern Standard Arabic
Abdullah Salem Khered, Youcef Benkhedda and Riza Batista-Navarro 50

Web-Based Corpus Compilation of the Emirati Arabic Dialect
Yousra A. El-Ghawi 63

Evaluating Calibration of Arabic Pre-trained Language Models on Dialectal Text
Ali Al-Laith and Rachida Kebdani 68

Empirical Evaluation of Pre-trained Language Models for Summarizing Moroccan Darija News Articles
Azzedine Aftiss, Salima Lamsiyah, Christoph Schommer and Said Ouatik El Alaoui 77

Dialect2SQL: A Novel Text-to-SQL Dataset for Arabic Dialects with a Focus on Moroccan Darija
Salmane Chafik, Saad Ezzini and Ismail Berrada 86

AraSim: Optimizing Arabic Dialect Translation in Children's Literature with LLMs and Similarity Scores
Alaa Hassan Bouomar and Noorhan Abbas 93

Navigating Dialectal Bias and Ethical Complexities in Levantine Arabic Hate Speech Detection
Ahmed Haj Ahmed, Rui-Jie Yew, Xerxes Minocher and Suresh Venkatasubramanian 103