

# **5th Workshop on Open-Source Arabic Corpora and Processing Tools with a Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection (OSACT 2022)**

Held at LREC 2022

Marseille, France  
20 June 2022

## **Editors:**

**Hend Al-Khalifa  
Tamer Elsayed  
Hamdy Mubarak**

**Abdulmohsen Al-Thubaity  
Walid Magdy  
Kareem Darwish**

ISBN: 978-1-7138-6131-7

**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 European Language Resources Association (ELRA)  
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-NC. (<https://creativecommons.org/licenses/by-nc/4.0/>)

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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>TURJUMAN: A Public Toolkit for Neural Arabic Machine Translation</i> El Moatez Billah Nagoudi, AbdelRahim Elmadany and Muhammad Abdul-Mageed .....	1
<i>Detecting Users Prone to Spread Fake News on Arabic Twitter</i> Zien Sheikh Ali, Abdulaziz Al-Ali and Tamer Elsayed .....	12
<i>AraSAS: The Open Source Arabic Semantic Tagger</i> Mahmoud El-Haj, Elvis de Souza, Nouran Khallaf, Paul Rayson and Nizar Habash .....	23
<i>AraNPCC: The Arabic Newspaper COVID-19 Corpus</i> Abdulmohsen Al-Thubaity, Sakhar Alkhereyf and Alia O. Bahanshal .....	32
<i>Pre-trained Models or Feature Engineering: The Case of Dialectal Arabic</i> Kathrein Abu Kwaik, Stergios Chatzikyriakidis and Simon Dobnik .....	41
<i>A Context-free Arabic Emoji Sentiment Lexicon (CF-Arab-ESL)</i> Shatha Ali A. Hakami, Robert Hendley and Phillip Smith .....	51
<i>Sa'7r: A Saudi Dialect Irony Dataset</i> Halah AlMazrua, Najla AlHazzani, Amaal AlDawod, Lama AlAwlaqi, Noura AlReshoudi, Hend Al-Khalifa and Luluh AlDhubayi .....	60
<i>Classifying Arabic Crisis Tweets using Data Selection and Pre-trained Language Models</i> Alaa Alharbi and Mark Lee .....	71
<i>Qur'an QA 2022: Overview of The First Shared Task on Question Answering over the Holy Qur'an</i> Rana Malhas, Watheq Mansour and Tamer Elsayed .....	79
<i>DTW at Qur'an QA 2022: Utilising Transfer Learning with Transformers for Question Answering in a Low-resource Domain</i> Damith Premasiri, Tharindu Ranasinghe, Wajdi Zaghouni and Ruslan Mitkov .....	88
<i>eRock at Qur'an QA 2022: Contemporary Deep Neural Networks for Qur'an based Reading Comprehension Question Answers</i> Esha Aftab and Muhammad Kamran Malik .....	96
<i>GOF at Qur'an QA 2022: Towards an Efficient Question Answering For The Holy Qu'ran In The Arabic Language Using Deep Learning-Based Approach</i> Ali Mostafa and Omar Mohamed .....	104
<i>LARSA22 at Qur'an QA 2022: Text-to-Text Transformer for Finding Answers to Questions from Qur'an</i> Youssef MELLAH, Ibtissam Touahri, Zakaria Kaddari, Zakaria Haja, Jamal Berrich and Toumi Bouchentouf .....	112
<i>LK2022 at Qur'an QA 2022: Simple Transformers Model for Finding Answers to Questions from Qur'an</i> Abdullah Alsaleh, Saud Althabiti, Ibtisam Alshammari, Sarah Alnefaie, Sanaa Alowaidi, Alaa Alsager, Eric Atwell, Abdulrahman Altahhan and Mohammad Alsalka .....	120

<i>niksss at Qur'an QA 2022: A Heavily Optimized BERT Based Model for Answering Questions from the Holy Qu'ran</i> Nikhil Singh .....	126
<i>QQATeam at Qur'an QA 2022: Fine-Tuning Arabic QA Models for Qur'an QA Task</i> Basem Ahmed, Motaz Saad and Eshrag A. Refaee, .....	130
<i>SMASH at Qur'an QA 2022: Creating Better Faithful Data Splits for Low-resourced Question Answering Scenarios</i> Amr Keleg and Walid Magdy .....	136
<i>Stars at Qur'an QA 2022: Building Automatic Extractive Question Answering Systems for the Holy Qur'an with Transformer Models and Releasing a New Dataset</i> Ahmed Sleem, Eman Mohammed lotfy Elrefai, Marwa Mohammed Matar and Haq Nawaz ...	146
<i>TCE at Qur'an QA 2022: Arabic Language Question Answering Over Holy Qur'an Using a Post-Processed Ensemble of BERT-based Models</i> Mohamemd Elkomy and Amany M. Sarhan .....	154
<i>Overview of OSACT5 Shared Task on Arabic Offensive Language and Hate Speech Detection</i> Hamdy Mubarak, Hend Al-Khalifa and Abdulmohsen Al-Thubaity .....	162
<i>GOF at Arabic Hate Speech 2022: Breaking The Loss Function Convention For Data-Imbalanced Arabic Offensive Text Detection</i> Ali Mostafa, Omar Mohamed and Ali Ashraf .....	167
<i>iCompass at Arabic Hate Speech 2022: Detect Hate Speech Using QRNN and Transformers</i> Mohamed Aziz Benessir, Malek Rhouma, Hatem Haddad and Chayma Fourati .....	176
<i>UPV at the Arabic Hate Speech 2022 Shared Task: Offensive Language and Hate Speech Detection using Transformers and Ensemble Models</i> Angel Felipe Magnossão de Paula, Paolo Rosso, Imene Bensalem and Wajdi Zaghouni .....	181
<i>Meta AI at Arabic Hate Speech 2022: MultiTask Learning with Self-Correction for Hate Speech Classification</i> Badr AlKhamissi and Mona Diab .....	186
<i>CHILLAX - at Arabic Hate Speech 2022: A Hybrid Machine Learning and Transformers based Model to Detect Arabic Offensive and Hate Speech</i> Kirolos Makram, Kirolos George Nessim, Malak Emad Abd-Almalak, Shady Zekry Roshdy, Seif Hesham Salem, Fady Fayek Thabet and Ensaf Hussien Mohamed .....	194
<i>AlexU-AIC at Arabic Hate Speech 2022: Contrast to Classify</i> Ahmad Shapiro, Ayman Khalafallah and Marwan Torki .....	200
<i>GUCT at Arabic Hate Speech 2022: Towards a Better Isotropy for Hatespeech Detection</i> Nehal Elkaref and Mervat Abu-Elkheir .....	209
<i>aiXplain at Arabic Hate Speech 2022: An Ensemble Based Approach to Detecting Offensive Tweets</i> Salaheddin Alzubi, Thiago Castro Ferreira, Lucas Pavanelli and Mohamed Al-Badrashiny ...	214