

NSURL 2021

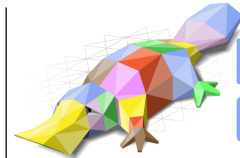
**Proceedings of the 4th International Conference on  
Natural Language and Speech Processing: Workshop on NLP  
Solutions for Under Resourced Languages**

November 14, 2021 (virtual)

موضوع



**DataScientia**  
Unitas per Varietatem



KNOW  
DIVE



**Tshwane University  
of Technology**

*We empower people*



©2021 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
acl@aclweb.org

ISBN 978-1-955917-19-3

<http://nsurl.org/>

## Introduction

Welcome to NSURL2021, the second International workshop on NLP Solutions for Under Resourced Languages (NSURL 2021) co-located with ICNLSP 2021, held on November 14th 2021.

NSURL is an opportunity and a forum for researchers and students to exchange ideas and discuss research and trends in the field of Natural Language Processing and Speech Processing by organizing shared tasks for solving NLP problems. This year, the task was on Semantic Relation Extraction in Persian. 10 papers have been submitted to NSURL 2021. 6 of them have been accepted. All the papers have been presented orally.

The attendance benefited from the keynotes presented at ICNLSP 2021. The first keynote was held by Dr. Ahmed Abdelali -QCRI- who presented his talk about understanding Arabic transformer Models. The second keynote entitled Figurative language analysis was given by PD Dr. Valia Kordoni -Humboldt University- followed by Dr. Hussein Al-Natsheh -Beyond limits- who gave interesting thoughts on AI technology commercialization and how to move from research to product innovation. The last talk was presented by Dr. Kareem Darwish -Aixplain- on one of the challenged topics which is Arabic diacritic recovery under the title Bring All Your Features: Arabic Diacritic Recovery Using a Feature-Rich Recurrent Neural Model.

We would like to acknowledge the support provided by University of Trento, and KnowDive group (University of Trento), and Datascentia (University of Trento). We would like also to express our gratitude to the organizing and the program committees for the hard and valuable contributions.

Abed Alhakim Freihat and Mourad Abbas

**Organizers:**

*Chair:* Dr. Abed Alhakim Freihat

*Co-chair:* Dr. Mourad Abbas

**Program Committee:**

Dr. Mourad Abbas, HCLA, Algeria

Prof. Sunday Ojo, Tshwane University of Technology  
Pretoria, South Africa

Mohamed Lichouri, USTHB, Algeria

Ahmed AbuRa'ed, Universitat P. F. Barcelona, Spain

Prof Linda van Huyssteen

Tshwane University of Technology, Pretoria, South Africa

Dr. Abdulmohsen Althubaity, The National Center for Artificial  
Intelligence and Big Data (kacst), KSA

Dr. Mohammad Gharib, University of Florence, Italy

Dr. Violetta Cavalli-Sforza, Al Akhawayn University, Morocco

Dr. Hesham Faili, University of Tehran, Iran

Nasrin Taghizadeh, University of Tehran, Iran

Dr. Abdallah Abushmaes, Mawdoos Ltd, Jordan

Osama Hamed, University of Duisburg-Essen, Germany

Hadi Khalilia, Univeristy of Trento, Italy

Nandu C Nair, Univeristy of Trento, Italy

**Organizing Committee:**

Hadi Khalilia, Univeristy of Trento

Nandu C Nair, Univeristy of Trento

## Table of Contents

<b>IDPL-PFOD: An Image Dataset of Printed Farsi Text for OCR Research</b> . . . . .	1
<i>Fatemeh sadat Hosseini, Shima Kashef, Elham Shabaninia, Hossein Nezamabadi-pour</i>	
<b>Improving Persian Relation Extraction Models by Data Augmentation</b> . . . . .	11
<i>Moein Salimi Sartakhti, Romina Etezadi and Mehrnoush Shamsfard</i>	
<b>NSURL-2021 Task 1: Semantic Relation Extraction in Persian</b> . . . . .	17
<i>Nasrin Taghizadeh, Ali Ebrahimi and Heshaam Faili</i>	
<b>PerSpellData: An Exhaustive Parallel Spell Dataset For Persian</b> . . . . .	24
<i>Romina Oji, Nasrin Taghizadeh and Heshaam Faili</i>	
<b>Improving Pre-Trained Language Model for Relation Extraction Using Syntactic Information in Persian</b> . . .	31
<i>Mohammad Mahdi Jafari, Somayyeh Behmanesh, Alireza Talebpour and Ali Nadian Ghomsheh</i>	
<b>The Dimensions of Lexical Semantic Resource Quality</b> . . . . .	38
<i>Hadi Khalilia, Abed Alhakim Freihat and Fausto Giunchiglia</i>	