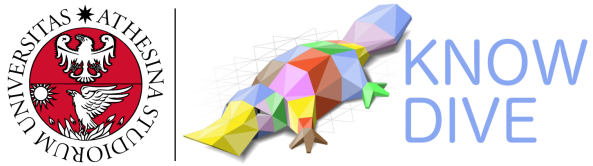
ICNLSP 2021

Proceedings of the 4th International Conference on Natural Language and Speech Processing

12–13 November, 2021 (virtual)




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 978-1-955917-18-6

https://www.icnlsp.org/

Introduction

Welcome to the fourth International Conference on Natural Language and Speech Processing (ICNLSP 2021), held online on November 12th, 13th 2021. ICNLSP is an opportunity and a forum for researchers, students, and industrials to exchange ideas and discuss research and trends in the field of Natural Language Processing. Indeed, many topics were discussed through the interesting works presented during the two days of the conference: speech recognition, machine translation, text summarization, sentiment analysis, natural language understanding, language resources, etc.

The accepted papers are of good quality thanks to the high-quality level of the reviews done by the program committee members who decided to accept 35 papers (long and short ones).

ICNLSP 2021, by including the second NSURL workshop, aims to draw the attention of researchers to provide solutions and resources for under resourced languages, by organizing shared tasks/ competitions for solving NLP problems. This year, the task was on Semantic Relation Extraction in Persian which attracted a number of contributions, 6 of them were accepted and presented in the workshop on November 14th, 2021.

We had the honor of having high-standard speakers with us, who gave valuable talks, starting 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 Datascientia (University of Trento). We would like also to express our gratitude to the organizing and the program committees for the hard and valuable contributions.

Mourad Abbas and Abed Alhakim Freihat

Organizers:

General Chair: Dr. Mourad Abbas
Chair: Dr. Abed Alhakim Freihat

Program Chair: Dr. Abed Alhakim Freihat

Publicity Chair: Dr. Mohamed Lichouri

Program Committee:

Mourad Abbas, HCLA, Algeria
Ahmed Abdelali, QCRI, Qatar
Mohamed Afify, Microsoft, Egypt
Messaoud Bengherabi, CDTA, Algeria
Djamel Bouchaffra, CDTA, Algeria
Fayssal Bouarourou, University of Strasbourg, France
Markus Brückl, TU Berlin, Germany
Hadda Cherroun, Amar Telidji University, Algeria
Gérard Chollet, CNRS, France
Kareem Darwish, QCRI, Qatar
Najim Dehak, Johns Hopkins University, USA
Mohamed Elfeky, Google Inc., USA
Ashraf Elnagar, University of Sharjah, UAE
Abed Alhakim Freihat, University of Trento, Italy
Nada Ghneim, Syrian Virtual University, Syria
Neil Glackin, Intelligent Voice, UK
Ahmed Guessoum, USTHB, Algeria
Mahmoud Gzawi, university of Lyon 2, France
Valia Kordoni, Humboldt University, Germany
Tomi Kinnunen, University of Eastern Finland, Finland
Eric Laporte, UPEM, France
Shang-Wen Li, Facebook AI., USA
Georges Linarès, University of Avignon, France
Shervin Malmasi, Harvard University, USA
Lluis Marquez, Amazon, Spain
Mhamed Mataoui, EMP, Algeria
Mohammed Mediani, University of Adrar, Algeria
Fatiha Merazka, USTHB, Algeria
Hamdy Mubarak, QCRI, Qatar
Preslav Nakov, QCRI, Qatar
Alexis Neme, UPEM, France
Axel Roebel, IRCAM, France
Younes Samih, Universität Düsseldorf, Germany

Hassan Satori, Sidi Mohammed Ben Abbdallah University, Morocco
Tim Schlippe, Silicon Surfer, Germany
Khaled Shaalan, The British University in Dubai, UAE
Otakar Smrz, Džám-e Džam Language Institute, Czech Republic
Rudolph Sock, University of Strasbourg, France
Irina Temnikova, QCRI, Qatar
Jan Trmal, Johns Hopkins University, USA
Stephan Vogel, QCRI, Qatar
Fayçal Ykhlef, CDTA, Algeria
Hasna Zaouali, University of Strasbourg, France

Additional Reviewers:

Hadi Khalilia, University of Trento, Italy
Mohamed Lichouri, USTHB, Algeria
Khaled Lounnas, USTHB, Algeria
Attia Nehar, University of Ziane Achour, Algeria
Slimane Bellaouar, University of Ghardaia, Algeria.

Organizing Committee:

Hadi Khalilia, University of Trento
Khaled Lounnas, USTHB, Algeria
Nandu C Nair, University of Trento

Invited Speakers:

Dr. Ahmed Abdelali, QCRI, Qatar
PD Dr. Valia Kordoni, Humboldt-Universität zu Berlin, Germany
Dr. Hussein Al-Natsheh, Beyond Limits.
Dr. Kareem Darwish, AiXplain.

Invited Talks

Understanding Arabic Transformer Models

Dr. Ahmed Abdelali

The success of pre-trained transformer models trained on Arabic and its dialects have gained more attention in the last few years . They were able to set and achieve new state of the art performance and accuracy in numerous downstream NLP tasks. Despite such popularity, no evaluation to compare the internal representations has been conducted. In this work we present deep comparison for these pre-trained Arabic models beyond the data used for the training or detailed architecture. We present an in-depth analysis for the layers and neurons for these models. The evaluation is done using three intrinsic tasks: two morphological tagging tasks based on MSA (modern standard Arabic) and dialectal Arabic and a dialectal identification task.

Figurative Language Analysis

PD Dr. Valia Kordoni

This talk focuses on figurative language analysis in multi-genre data. While metaphor has been tackled in Natural Language Processing before, the focus has never simultaneously been on the analysis of multi-genre and heterogeneous texts.

AI Technology Commercialization: From Research to Product Innovation

Dr. Hussein Al-Natsheh

The number of researchers in the NLP research community, and the AI research at large, is increasing as well as the funding from both the public and private sectors. However, not enough of these invented technologies are applied in solving real-life problems. In this keynote, we will shed the light on this challenge and how we can turn it into an opportunity that can motivate investing in more research both applied and scientific. This topic touches many areas that we will present and link to in the session including open innovation, open-source, open data, technology licensing, product innovation, marketing and pricing models, investment, team building, and MLOps. We will also provide some examples where we have successfully turned research-level technology into successful and scalable products.

Bring All Your Features: Arabic Diacritic Recovery Using a Feature-Rich Recurrent Neural Model

Dr. Kareem Darwish

Diacritics (short vowels) are typically omitted when writing Arabic text, and readers have to reintroduce them to correctly pronounce words. There are two types of Arabic diacritics: the first are core-word diacritics (CW), which specify the lexical selection, and the second are case endings (CE), which typically appear at the end of word stems and generally specify their syntactic roles. Recovering CEs is significantly harder than recovering core-word diacritics due to inter-word dependencies, which are often distant. The presentation shows the use of a feature-rich recurrent neural network model that uses a variety of linguistic and surface-level features to recover both core word diacritics and case endings. The model surpasses all previous state-of-the-art systems with a CW error rate of 2.86% and a CE error rate (CEER) of 3.7%, which is 61% lower than any state-of-the-art system. When combining diacritized word cores with case endings, the resultant word error rate is 6.0%. This highlights the effectiveness of feature engineering for such deep neural models.

Table of Contents