Welcome to ICNLSP 2019, the third edition of the International Conference on Natural Language and Speech Processing, held on September 12th, 13th 2019, and hosted at the university of Trento in Italy. ICNLSP 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.

The 46 papers submitted to ICNLSP 2019 have been reviewed by 3 reviewers. The program committee decided to accept 20 of them with an acceptance rate of 43 %. The papers will be presented orally. They cover various topics dealing with both speech and text: building resources, text summarization, spoken language understanding, etc.

The program includes also two keynotes. The first one, entitled "Detecting the fake news before they were even written", will be presented by Dr. Preslav Nakov from Qatar Computing Research Institute (QCRI), Qatar. The second keynote "One world - seven thousand languages" will be presented by Prof. Fausto Giunchiglia from University of Trento, Italy.

The conference is preceded by the workshop on NLP Solutions for Under Resourced Languages (NSURL). The workshop is intended as a forum for solving NLP problems for low-resourced languages.

We would like to acknowledge the support provided by University of Trento and DataScientia. We would like also to express our gratitude to the organizing and the program committees for the hard and valuable contributions.

We hope that ICNLSP 2019 will be highly positive scientific event. We wish a happy time for all the participants.

Mourad Abbas, and Abed Alhakim Freihat
Trento, September 2019
Organizers:

General Chair: Mourad Abbas
Chair: Abed Alhakim Freihat

Program Committee:

Mourad Abbas, CRSTDLA, Algeria
Ahmed Abdelali, QCRI, Qatar
Mohamed Afify, Microsoft, Egypt
Mansour Alghamdi, KACST, Saudi Arabia
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
Najim Dehak, Johns Hopkins University, USA
Mohamed Elfeky, Google Inc., USA
Mahmoud El-Haj, Lancaster University, UK
Abed Alhakim Freihat, University of Trento, Italy
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, Apple Inc., USA
Georges Linarès, University of Avignon, France
Walid Magdy, University of Edinburgh, UK
Shervin Malmasi, Harvard University, USA
Lluis Marquez, Amazon, Spain
Mohammed Mediani, University of Adrar, Algeria
Fatiha Merazka, USTHB, Algeria
Hamdy Mubarak, QCRI, Qatar
Preslav Nakov, QCRI, Qatar
Alexis Neme, UPEM, France
Mourad Ouzzani, QCRI, Qatar
Ahmed Rafea, American University in Cairo, Egypt
Abdelmounaam Rezgui, New Mexico Tech, USA
Younes Samih, Universität Düsseldorf, Germany
Violetta Cavalli Sforza, Al Akhawayn 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, Professor, University of Strasbourg, France
Irina Temnikova, QCRI, Qatar
Jan Trmal, Johns Hopkins University, USA
Stephan Vogel, QCRI, Qatar
Marcos Zampieri, University of Wolverhampton, UK
Hasna Zaouali, University of Strasbourg, France

Additional reviewers:
Mohamed Azzaz, EMP, Algeria
Riadh Belkebir, USTHB, Algeria
Mohamed Lichouri, CRSTDLA, Algeria
Mhamed Mataoui, EMP, Algeria
Attia Nehar, Ziane Achour University, Algeria
Said Sadoudi, EMP, Algeria
Fayçal Ykhlef, CDTA, Algeria

Organizing committee:
Gabor Bella, University of Trento.
Mattia Fumagalli, University of Trento.
Nandu C Naird, University of Trento.
Olha Vozna, University of Trento.
Mohammad Gharib, University of Florence.
Moaz Reyad, University of Genoa.
Osama Hamed, University of Duisburg-Essen.
Mohamed Lichouri, CRSTDLA.

Invited Speakers:
Prof. Fausto Giunchiglia, University of Trento, Italy.
Dr. Preslav Nakov, Qatar Computing Research Institute (QCRI), Qatar.
Invited Talks

Detecting the "Fake News" before they were even written
Preslav Nakov

Given the recent proliferation of disinformation online, there has been also growing research interest in automatically debunking rumors, false claims, and "fake news". A number of fact-checking initiatives have been launched so far, both manual and automatic, but the whole enterprise remains in a state of crisis: by the time a claim is finally fact-checked, it could have reached millions of users, and the harm caused could hardly be undone. An arguably more promising direction is to focus on fact-checking entire news outlets, which can be done in advance. Then, we could fact-check the news before they were even written: by checking how trustworthy the outlets that published them are.

We will show how we do this in the Tanbih news aggregator (http://www.tanbih.org/), which makes users aware of what they are reading. In particular, we develop media profiles that show the general factuality of reporting, the degree of propagandistic content, hyper-partisanship, leading political ideology, general frame of reporting, stance with respect to various claims and topics, as well as audience reach and audience bias in social media.

One world - seven thousand languages
Fausto Giunchiglia

We present a large scale multilingual lexical resource, the Universal Knowledge Core (UKC), which is organized like a Wordnet with, however, a major design difference. In the UKC, the meaning of words is represented not only with synsets, but also using language independent concepts which cluster together the synsets which, in different languages, codify the same meaning. In the UKC, it is concepts and not synsets, as it is the case in the Wordnets, which are connected in a semantic network. The use of language independent concepts allows for the native integrability, analysis and use of any number of languages, with important applications in, e.g., multilingual language processing, reasoning (as needed, for instance, in data and knowledge integration) and image understanding.
# Table of Contents

Twitter Bot Detection using Diversity Measures .......................... 1  
*Dijana Kosmajac and Vlado Keselj*

Aligning the IndoWordNet with the Princeton WordNet .................. 9  
*Nandu Chandran Nair, Rajendran Sankara Velayuthan and Khuyagbaatar Batsuren*

Automatic Data-Driven Approaches for Evaluating the Phonemic Verbal Fluency Task with Healthy Adults ............................................. 17  
*Hali Lindsay, Nicklas Linz, Johannes Troeger and Jan Alexandersson*

Automatic Detection and Classification of Argument Components using Multi-task Deep Neural Network .................................................. 25  
*Jean-Christophe Mensonides, Sébastien Harispe, Jacky Montmain and Véronique Thirieu*

Multi Sense Embeddings from Topic Models .................................. 34  
*Shobhit Jain, Sravan Babu Bodapati, Ramesh Nallapati and Anima Anandkumar*

Automatic Arabic Text Summarization Based on Fuzzy Logic ............ 42  
*Lamees Al Qassem, Di Wang, Hassan Barada, Ahmad Al-Rubaie and Nawaf Almoosa*

An Arabic Multi-Domain Spoken Language Understanding System ........ 49  
*Mohamed Lichouri, Mourad Abbas, Rachida Djeradi and Amar Djeradi*

Building a Speech Corpus based on Arabic Podcasts for Language and Dialect Identification .................. 54  
*Khaled Lounnas, Mourad Abbas and Mohamed Lichouri*

Automatic Text Tagging of Arabic News Articles Using Ensemble Deep Learning Models .......................... 59  
*Ashraf Elnagar, Omar Einea and Ridhwan Al-Debsi*

A Probabilistic Approach for Confidence Scoring in Speech Recognition ........ 67  
*Punnoose Kuriakose*

A Crowdsourcing-based Approach for Speech Corpus Transcription Case of Arabic Algerian Dialects .......... 75  
*Ilyes Zine, Mohamed Cherif Zeghad, Soumia Bougrine and Hadda Cherroun*

Sample Size in Arabic Authorship Verification ................................ 84  
*Hossam Ahmed*

A folksonomy-based approach for profiling human perception on word similarity .......................... 92  
*Guani Wu and Ker-Chau Li*

Automatic Diacritization as Prerequisite Towards the Automatic Generation of Arabic Lexical Recognition Tests ............................................. 100  
*Osama Hamed*

Expanding English and Chinese Dictionaries by Wikipedia Titles ........ 107  
*Wei-Ting Chen, Yu-Te Wang and Chuan-Jie Lin*

Production of Voicing Contrast in Children with Cochlear Implants ........ 114  
*Georgia Koupka*

SumSAT: Hybrid Arabic Text Summarization based on symbolic and numerical Approaches ........ 120  
*Mohamed Amine Cheragui and Said Moulay Lakhdar*

Speech Coding Combining Chaos Encryption and Error Recovery for G.722.2 Codec .......................... 128
Messaouda Boumaraf and Fatiha Merazka
Compositional pre-training for neural semantic parsing ........................................ 135
Amir Ziai
Encoding Position Improves Recurrent Neural Text Summarizers ..................... 142
Apostolos Karanikolos and Ioannis Refanidis