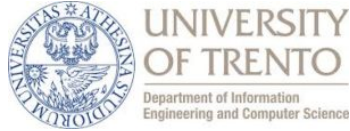


ICNLSP 2019

**Proceedings of the First International Workshop on
NLP Solutions for Under Resourced Languages (NSURL 2019)
co-located with ICNLSP 2019**

11–12 September, 2019
University of Trento
Trento, Italy



©2019 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 978-1-952148-53-8

Introduction

Welcome to NSURL2019, the First International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019) co-located with ICNLSP 2019, held on September 11th, 12th 2019, at the University of Trento in Italy. 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. 26 papers have been submitted to NSURL 2019. 19 of them have been accepted. All the papers have been presented orally. The workshop, indeed, has been an interesting forum for solving NLP problems for low-resourced languages.

The attendance benefited from the two keynotes presented at ICNLSP 2019. The first one, entitled "Detecting the fake news before they were even written", presented by Dr. Preslav Nakov from Qatar Computing Research Institute (QCRI), Qatar. The second keynote "One world - seven thousand languages" presented by Prof. Fausto Giunchiglia from University of Trento, Italy. We would like to acknowledge the support provided by University of Trento and Data-Scientia. 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

Trento, September 2019

Organizers:

Chair: Dr. Abed Alhakim Freihat

Co-Chair: Dr. Mourad Abbas

Program Committee:

Mourad Abbas, CRSTDLA, Algeria

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

Abdallah Abushmaes, Mawdoo3 Ltd, Jordan

Abdulmohsen Althubaity, The NCAIBD Research Center-KACST, KSA

Violetta Cavalli-Sforza, Al Akhawayn University, Morocco

Shumile Chabalala, Tshwane University of Technology, South Africa

Abdelrahim Elmadany, Uinversity of Jazan, KSA

Heshaam Faili, University of Tehran, Iran

Mohammad Gharib, University of Florence, Italy

Osama Hamed, University of Duisburg-Essen, Germany

Linda van Huyssteen, Tshwane University of Technology, South Africa

Gabriel Iwasokun, Federal university of Technology Akure, Nigeria

Mohamed Lichouri, CRSTDLA, Algeria

Itani P. Mandende, Tshwane University of Technology, South Africa

Charles Mann, Tshwane University of Technology, South Africa

Maredi I. Mphahlele, Tshwane University of Technology, South Africa

Nandu C Nair, Univeristy of Trento, Italy

Hussein Natsheh, Mawdoo3 Ltd, Jordan

Gabriel Ogunleye, Federal University,Oye-Ekiti, Nigeria

Sunday Ojo, Tshwane University of Technology, South Africa

O. Olugbara, Durban University of Technology, South Africa

Pius A. Owolawi, Tshwane University of Technology, South Africa

Agnieta B. Pretorius, Tshwane University of Technology, South Africa

Adeyanju Sosomi, University of Lagos, Nigeria

Nasrin Taghizadeh, University of Tehran, Iran

Etienne E. Van Wyk, Tshwane University of Technology, South Africa

Organizing committee:

Gabor Bella, Univeristy of Trento

Mattia Fumagalli, Univeristy of Trento

Nandu C Nair, Univeristy of Trento

Olha Vozna, University of Trento

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

NSURL-2019 Task 8: Semantic Question Similarity in Arabic	1
<i>Haitham Seelawi, Ahmad Mustafa, Hesham Al-Bataineh, Wael Farhan and Hussein T Al-Natsheh</i>	
NSURL-2019 Task 7: Named Entity Recognition for Farsi	9
<i>Nasrin Taghizadeh and Hesham Faili</i>	
Yorùbá Gender Recognition from Speech using Attention-based BiLSTM	16
<i>Ibukunola Abosede Modupe, Tshephisho Joseph Sefara and Ojo Sunday</i>	
MorphoBERT: a Persian NER System with BERT and Morphological Analysis	23
<i>Mahdi Mohseni and Amirhossein Tebbifakhr</i>	
AtyNegar at NSURL-2019 Task 8: Semantic Question Similarity in Arabic	31
<i>Atieh Sharifi, Hossein Hassanpoor and Najmeh Zare Maduyieh</i>	
Beheshti-NER: Persian named entity recognition Using BERT	37
<i>Ehsan Taher, Seyed Abbas Hoseini and Mehrnoush Shamsfard</i>	
Arabic Dialogue Act Recognition for Textual Chatbot Systems	43
<i>Alaa Joukhadar, Huda Saghergy, Leen Kweider and Nada Ghneim</i>	
Tha3aroon at NSURL-2019 Task 8: Semantic Question Similarity in Arabic	50
<i>Ali Fadel, Ibraheem Tuffaha and Mahmoud Al-Ayyoub</i>	
Motivations, challenges, and perspectives for the development of an Automatic Speech Recognition System for the under-resourced Ngiemboon Language	58
<i>Patrice Yemmene and Laurent Besacier</i>	
NITK-IT_NLP@NSURL2019: Transfer Learning based POS Tagger for Under Resourced Bhoj-puri and Magahi Language	67
<i>Anand Kumar M</i>	
The_Illiterati: Part-of-Speech Tagging for Magahi and Bhojpurii without even knowing the alphabet	72
<i>Thomas Proisl, Peter Uhrig, Andreas Blombach, Natalie Dykes, Philipp Heinrich, Besim Kabashi and Sefora Mammarella</i>	
ST NSURL 2019 Shared Task: Semantic Question Similarity in Arabic	79
<i>Mohamed Lichouri, Mourad Abbas, Besma Benaziz and Abed Alhakim Freihat</i>	
Statistical Machine Translation between Myanmar (Burmese) and Dawei (Tavoyan)	84
<i>Thazin Myint Oo, Ye Kyaw Thu, Khin Mar Soe and Thepchai Supnithi</i>	
String Similarity Measures for Myanmar Language (Burmese)	93
<i>Khaing Hsu Wai, Ye Kyaw Thu, Hnin Aye Thant, Swe Zin Moe and Thepchai Supnithi</i>	
An Inferential Phonological Connectionist Approach to the perception of Assimilated-English Connected Speech	102
<i>Hiba Zaidi</i>	
Improving NER Models by exploiting Named Entity Gazetteer as External Knowledge	106
<i>Atefeh Zafarian and Habibollah Asghari</i>	
The Inception Team at NSURL-2019 Task 8: Semantic Question Similarity in Arabic	111
<i>Hana Al-Theiabat and Aisha Al-Sadi</i>	

Hidden Markov-based Part-of-Speech Tagger for Igbo Resource-Scarce African Language	117
<i>Ihenaetu Olamma, Michael Kingsley and Sunday Ojo</i>	
Building Ontology for Yorùbá Language	123
<i>Theresa Okediya, Ibukun Afolabi, Olamma Iheanetu and Sunday Ojo</i>	