

NAACL-HLT 2021

**The 2021 Conference
of the North American Chapter
of the Association for Computational Linguistics:
Human Language Technologies**

Proceedings of the Student Research Workshop

June 6 - 11, 2021

©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 978-1-954085-50-3

Introduction

Welcome to the NAACL 2021 Student Research Workshop!

The NAACL 2021 Student Research Workshop (SRW) is a forum for student researchers in computational linguistics and natural language processing. The workshop provides a unique opportunity for student participants to present their work and receive valuable feedback from the international research community as well as from faculty mentors.

Following the tradition of the previous student research workshops, we have two tracks: research papers and thesis proposals. The research paper track is a venue for Ph.D. students, Masters students, and advanced undergraduates to describe completed work or work-in-progress along with preliminary results. The thesis proposal track is offered for advanced Masters and Ph.D. students who have decided on a thesis topic and are interested in feedback on their proposal and ideas about future directions for their work.

This year, we received 50 submissions in total. We accepted 22 papers, with an acceptance rate of 44

Mentoring is at the heart of the SRW. In keeping with previous years, we had a pre-submission mentoring program before the submission deadline. A total of 12 papers participated in the pre-submission mentoring program. This program offered students the opportunity to receive comments from an experienced researcher to improve the writing style and presentation of their submissions.

We are deeply grateful to our sponsor, the National Science Foundation. We thank our program committee members for their careful reviews of each paper and all of our mentors for donating their time to provide feedback to our student authors. Thank you to our faculty advisors, Nanyun Peng and Yu Su, for their essential advice and guidance, and to the NAACL 2021 organizing committee for their support. Finally, thank you to our student participants!

Organizers

Esin Durmus - Cornell University
Vivek Gupta - University of Utah
Nelson Liu - Stanford University

Faculty Advisors

Nanyun Peng - University of California, Los Angeles
Yu Su - The Ohio State University

Pre-submission Mentors

Mohit Bansal - University of North Carolina, Chapel Hill
Valerio Basile - University of Turin
Eduardo Blanco - University of North Texas
Dallas Card - Stanford University
Kai-Wei Chang - University of California, Los Angeles
Eunsol Choi - University of Texas, Austin
Lucia Donatelli - Saarland University
Greg Durrett - University of Texas, Austin
Yansong Feng - Peking University
Matt Gardner - Allen Institute for Artificial Intelligence
Luke Holman - University of North Texas
Robin Jia - Facebook AI Research / University of Southern California
Debanjan Mahata - Bloomberg
Vincent Ng - University of Texas, Dallas
Barbara Plank - IT University of Copenhagen
Sai Krishna Rallabandi - Carnegie Mellon University
Melissa Roemmele - SDL
Masoud Rouhizadeh - Johns Hopkins University
Sebastian Ruder - DeepMind
Roy Schwartz - Hebrew University of Jerusalem
Vered Shwartz - Allen Institute for Artificial Intelligence
Maneesh Singh - Verisk Inc.
Amanda Stent - Bloomberg
Swabha Swayamdipta - Allen Institute for Artificial Intelligence
Jesse Thomason - Amazon / University of Southern California
Aline Villavicencio - Federal University of Rio Grande do Sul
Bonnie Webber - University of Edinburgh
Arkaitz Zubiaga - Queen Mary University of London

Program Committee

Piush Aggarwal
Afroz Ahamad
Chaitanya Ahuja
Abeer AL-Dayel
Sedeeq Al-khazraji

Miguel A. Alonso
Rami Aly
Bharat Ram Ambati
Calude Andreea
Maria Antoniak
Kristjan Arumae
Ehsaneddin Asgari
Vidhisha Balachandran
Loïc Barrault
Valerio Basile
Roberto Basili
Rachel Bawden
Eduardo Blanco
Rishi Bommasani
Marine Carpuat
Arlene Casey
Arun Chaganty
lisa andreevna chalaguine
Mingda Chen
Xinchi Chen
Jifan Chen
Leshem Choshen
Siddharth Dalmia
Samvit Dammalapati
Alok Debnath
Louise Deléger
Chris DeVelder
Flavio Di Palo
Anne Dirkson
Radina Dobрева
Zi-Yi Dou
Micha Elsner
Denis Emelin
Carlos Escolano
Tina Fang
Murhaf Fares
Amir Feder
Yansong Feng
Jared Fernandez
Dayne Freitag
Yoshinari Fujinuma
Saadia Gabriel
Diana Galvan-Sosa
Marcos Garcia
Matt Gardner
Dan Goldwasser
Sarah Gupta
Abhinav Gupta
Vivek Gupta
Hardy Hardy
Mareike Hartmann

Junxian He
Christopher Homan
saghar Hosseini
Mohammad Javad Hosseini
Nina Hosseini-Kivanani
Phu Mon Htut
Jeff Jacobs
Aaron Jaech
Glorianna Jagfeld
Labiba Jahan
Jyoti Jha
Zhengbao Jiang
Chen Junjie
Jad Kabbara
Tomoyuki Kajiwara
Ehsan Kamaloo
Zara Kancheva
Sudipta Kar
Alina Karakanta
Philipp Koehn
Taiwo Kolajo
Mamoru Komachi
Parisa Kordjamshidi
Mandy Korpusik
Kalpesh Krishna
Jonathan K. Kummerfeld
Kemal Kurniawan
Yash Kumar Lal
Alexandra Lavrentovich
Bowen Li
Yiyuan Li
Lei Li
Tao Li
Xiang Lorraine Li
Junwei Liang
Jasy Suet Yan Liew
Kevin Lin
Kevin Lin
Fangyu Liu
Debanjan Mahata
Valentin Malykh
Emma Manning
Pedro Henrique Martins
Bruno Martins
Puneet Mathur
Dheeraj Mekala
Omid Memarrast
Rui Meng
Antonio Valerio Miceli Barone
Sabrina J. Mielke
Zulfat Miftahudinov

Tsvetomila Mihaylova
Farjana Sultana Mim
Sewon Min
Koji Mineshima
Swaroop Mishra
Rohan Mishra
Amita Misra
Jesse Mu
Nora Muheim
Masaaki Nagata
Nihal V. Nayak
Dat Quoc Nguyen
Vincent Nguyen
Shinji Nishimoto
Pegah Nokhiz
Yasumasa Onoe
Naoki Otani
Endang Wahyu Pamungkas
Isabel Papadimitriou
Thiago Pardo
Chan Young Park
Roma Patel
Archita Pathak
Siyao Peng
Hai Pham
Adithya Pratapa
Ivaylo Radev
Sree Harsha Ramesh
Surangika Ranathunga
Vikas Raunak
Paul Rayson
Kirk Roberts
Alexander Robertson
Guy Rotman
Maria Ryskina
sepideh sadeghi
Younes Samih
Jainisha Sankhavara
Ryohei Sasano
Carolina Scarton
Michael Sejr Schlichtkrull
Sebastian Schuster
Giovanni Semeraro
Olga Seminck
Indira Sen
Gautam Kishore Shahi
Sina Sheikholeslami
A.B. Siddique
Kushagra Singh
Pradyumna Sinha
Sunayana Sitaram

Katira Soleymanzadeh
Thamar Solorio
Richard Sproat
Tejas Srinivasan
Marija Stanojevic
Alane Suhr
Jeniya Tabassum
Shabnam Tafreshi
Nicholas Tomlin
Trang Tran
Sowmya Vajjala
Alakananda Vempala
Rob Voigt
Teodora Vukovic
Eric Wallace
Bonnie Webber
John Wieting
Steven Wilson
Patrick Xia
Yumo Xu
Rongtian Ye
Da Yin
Michael Yoder
Xiaodong Yu
Meishan Zhang
Tianyi Zhang
Yanpeng Zhao
Arkaitz Zubiaga
Talha Çolakoğlu

Table of Contents

<i>Sampling and Filtering of Neural Machine Translation Distillation Data</i> Vilém Zouhar	1
<i>IceSum: An Icelandic Text Summarization Corpus</i> Jón Daðason, Hrafn Loftsson, Salome Sigurðardóttir and Þorsteinn Björnsson	9
<i>Negation typology and general representation models for cross-lingual zero-shot negation scope resolution in Russian, French, and Spanish.</i> Anastassia Shaitarova and Fabio Rinaldi	15
<i>Representations of Meaning in Neural Networks for NLP: a Thesis Proposal</i> Tomáš Musil	24
<i>Towards Layered Events and Schema Representations in Long Documents</i> Hans Ole Hatzel and Chris Biemann	32
<i>Parallel Text Alignment and Monolingual Parallel Corpus Creation from Philosophical Texts for Text Simplification</i> Stefan Paun	40
<i>Syntax-Based Attention Masking for Neural Machine Translation</i> Colin McDonald and David Chiang	47
<i>Multi-Modal Image Captioning for the Visually Impaired</i> Hiba Ahsan, Daivat Bhatt, Kaivan Shah and Nikita Bhalla	53
<i>Open-Domain Question Answering with Pre-Constructed Question Spaces</i> Jinfeng Xiao, Lidan Wang, Franck Dernoncourt, Trung Bui, Tong Sun and Jiawei Han	61
<i>A Sliding-Window Approach to Automatic Creation of Meeting Minutes</i> Jia Jin Koay, Alexander Roustai, Xiaojin Dai and Fei Liu	68
<i>Exploration and Discovery of the COVID-19 Literature through Semantic Visualization</i> Jingxuan Tu, Marc Verhagen, Brent Cochran and James Pustejovsky	76
<i>Shuffled-token Detection for Refining Pre-trained RoBERTa</i> Subhadarshi Panda, Anjali Agrawal, Jeewon Ha and Benjamin Bloch	88
<i>Morphology-Aware Meta-Embeddings for Tamil</i> Arjun Sai Krishnan and Seyoon Ragavan	94
<i>Seed Word Selection for Weakly-Supervised Text Classification with Unsupervised Error Estimation</i> Yiping Jin, Akshay Bhatia and Dittaya Wanvarie	112
<i>Multi-Task Learning of Generation and Classification for Emotion-Aware Dialogue Response Generation</i> Tatsuya Ide and Daisuke Kawahara	119
<i>Comparison of Grammatical Error Correction Using Back-Translation Models</i> Aomi Koyama, Kengo Hotate, Masahiro Kaneko and Mamoru Komachi	126
<i>Parallel sentences mining with transfer learning in an unsupervised setting</i> Yu Sun, Shaolin Zhu, Feng Yifan and Chenggang Mi	136

<i>Sentence Concatenation Approach to Data Augmentation for Neural Machine Translation</i> Seiichiro Kondo, Kengo Hotate, Tosho Hirasawa, Masahiro Kaneko and Mamoru Komachi . . .	143
<i>Emotion Classification in a Resource Constrained Language Using Transformer-based Approach</i> Avishek Das, Omar Sharif, Mohammed Moshiul Hoque and Iqbal H. Sarker	150
<i>Hie-BART: Document Summarization with Hierarchical BART</i> Kazuki Akiyama, Akihiro Tamura and Takashi Ninomiya	159

Conference Program

Saturday, June 5, 2021 UTC+0

16:00–17:20 SRW Session 1

Sampling and Filtering of Neural Machine Translation Distillation Data

Vilém Zouhar

Towards Multi-Modal Text-Image Retrieval to improve Human Reading

Florian Schneider, Özge Alacam, Xintong Wang and Chris Biemann

IceSum: An Icelandic Text Summarization Corpus

Jón Daðason, Hrafn Loftsson, Salome Sigurðardóttir and Þorsteinn Björnsson

Negation typology and general representation models for cross-lingual zero-shot negation scope resolution in Russian, French, and Spanish.

Anastassia Shaitarova and Fabio Rinaldi

Representations of Meaning in Neural Networks for NLP: a Thesis Proposal

Tomáš Musil

Towards Layered Events and Schema Representations in Long Documents

Hans Ole Hatzel and Chris Biemann

Parallel Text Alignment and Monolingual Parallel Corpus Creation from Philosophical Texts for Text Simplification

Stefan Paun

Saturday, June 5, 2021 UTC+0 (continued)

18:40–20:00 SRW Session 2

Syntax-Based Attention Masking for Neural Machine Translation

Colin McDonald and David Chiang

Multi-Modal Image Captioning for the Visually Impaired

Hiba Ahsan, Daivat Bhatt, Kaivan Shah and Nikita Bhalla

Open-Domain Question Answering with Pre-Constructed Question Spaces

Jinfeng Xiao, Lidan Wang, Franck Dernoncourt, Trung Bui, Tong Sun and Jiawei Han

A Character-and-Event-Based Framework for Detecting Literary Allusion in Internet Humor

Emily Shen

A Sliding-Window Approach to Automatic Creation of Meeting Minutes

Jia Jin Koay, Alexander Roustai, Xiaojin Dai and Fei Liu

Exploration and Discovery of the COVID-19 Literature through Semantic Visualization

Jingxuan Tu, Marc Verhagen, Brent Cochran and James Pustejovsky

Shuffled-token Detection for Refining Pre-trained RoBERTa

Subhadarshi Panda, Anjali Agrawal, Jeewon Ha and Benjamin Bloch

Morphology-Aware Meta-Embeddings for Tamil

Arjun Sai Krishnan and Seyoon Ragavan

Sunday, June 6, 2021 UTC+0

1:20–2:40 SRW Session 3

Seed Word Selection for Weakly-Supervised Text Classification with Unsupervised Error Estimation

Yiping Jin, Akshay Bhatia and Dittaya Wanvarie

Multi-Task Learning of Generation and Classification for Emotion-Aware Dialogue Response Generation

Tatsuya Ide and Daisuke Kawahara

Comparison of Grammatical Error Correction Using Back-Translation Models

Aomi Koyama, Kengo Hotate, Masahiro Kaneko and Mamoru Komachi

Parallel sentences mining with transfer learning in an unsupervised setting

Yu Sun, Shaolin Zhu, Feng Yifan and Chenggang Mi

Sentence Concatenation Approach to Data Augmentation for Neural Machine Translation

Seiichiro Kondo, Kengo Hotate, Tosho Hirasawa, Masahiro Kaneko and Mamoru Komachi

Emotion Classification in a Resource Constrained Language Using Transformer-based Approach

Avishek Das, Omar Sharif, Mohammed Moshiul Hoque and Iqbal H. Sarker

Hie-BART: Document Summarization with Hierarchical BART

Kazuki Akiyama, Akihiro Tamura and Takashi Ninomiya

