ALTA 2021

# Proceedings of the 19th Workshop of the Australasian Language Technology Association

8–10 December, 2021
Online

## Gold Sponsors:

## Silver Sponsors:

# Introduction

Welcome to the 19th edition of the Annual Workshop of the Australasian Language Technology Association (ALTA 2021). The purpose of ALTA is to promote language technology research and development in Australia and New Zealand. Every year ALTA hosts a workshop which is the key local forum for disseminating research in Natural Language Processing and Computational Linguistics, with presentations and posters from students, industry, and academic researchers. This year ALTA is hosted as a virtual workshop, due to the COVID-19 pandemic. In total we received 18 long, 8 short, and 2 abstract submissions and we accepted 15 long and 7 short papers to appear in the workshop, as well as the 2 extended abstracts. Each paper was reviewed by at least two members of the program committee, using a double-blind protocol. Great care was taken to avoid all conflicts of interest. Of all submissions, 25 were first-authored by students. We had submissions from a total of seven countries: Australia, New Zealand, Spain, France, Germany, Netherlands and United States. We are extremely grateful to the Programme Committee members for their time and their detailed and helpful comments and reviews. This year we had committee members from all over the globe including Australia, New Zealand, Japan, Sweden, Switzerland, United States and United Arab Emirates. Overall, there will be six oral presentation sessions and two virtual poster sessions. We also ran a shared task in Evidence Based Medicine (EBM) organised by Diego Mollá-Aliod (Macquarie University). Participants were invited to submit a system description paper, which are included in this volume with a light review by ALTA chairs. Finally, the workshop will feature keynotes from Barbara Plank (IT University of Copenhagen), Ben Hutchinson (Google) and Dirk Hovy (Bocconi University), following a tradition of bringing speakers from both academia and industry. ALTA 2021 is very grateful for the financial support generously offered by our sponsors. Without their contribution, the running of these events to bring together the NLP community of the Australasian region would have been a challenge. We would like to express sincere gratitude to our sponsors. We very much hope that you will have an enjoyable and inspiring time at ALTA 2021!

Afshin Rahimi, William Lane and Guido Zuccon

Brisbane, Australia

Dec 2021

**Organisers:**

*Program co-chairs:* Afshin Rahimi, William Lane and Guido Zuccon
*Program Execs:* Maria Kim, Sarvnaz Karimi, Meladel Mistica, Diego Mollá and Massimo Piccardi
*Shared task chair:* Diego Mollá

**Program Committee:**

Abeed Sarkar (Emory University)
Afshin Rahimi (The University of Queensland)
Andrea Schalley (Karlstad University)
Antonio Jimeno (RMIT University)
Diego Molla (Macquarie University)
Dominique Estival (Western Sydney University)
Fajri Koto (The University of Melbourne)
Gabriela Ferraro (CSIRO)
Gholamreza Haffari (Monash University)
Hamed Hassanzadeh (The Australian e-Health Research Centre, CSIRO)
Guido Zuccon (University of Queensland)
Hiyori Yoshikawa (Fujitsu Laboratories Ltd.)
Jennifer Biggs (Defence Science and Technology Group)
Jey Han Lau (University of Melbourne)
Karin Verspoor (RMIT University)
Kristin Stock (Massey University)
Lea Frermann (The University of Melbourne)
Lizhen Qu (Monash University)
Maria Kim (Defence Science and Technology Group)
Massimo Piccardi (University of Technology Sydney)
Mel Mistica (The University of Melbourne)
Nitin Indurkhya (The University of New South Wales)
Sarvnaz Karimi (CSIRO)
Scott Nowson (PwC Middle East)
Stephen Wan (CSIRO)
Sunghwan Mac Kim (Hara Research Lab)
Timothy Baldwin (The University of Melbourne)
Trevor Cohn (The University of Melbourne)
William Lane (Charles Darwin University)
Wray Lindsay Buntine (Monash University)
Xiang Dai (University of Copenhagen)

**Invited Speakers:**

Barbara Plank, IT University of Copenhagen
Ben Hutchinson, Google
Dirk Hovy, Bocconi University

# Invited Talks

**Barbara Plank: Tackling scarce and biased data for more inclusive Natural Language Processing**

Deep neural networks have revolutionised our field in recent years. Particularly contextualised representations obtained from large-scale language models have pushed frontiers. Despite of these advances, many challenges and research problems remain, due to the rich variability of language and a dreadful lack and bias in resources. In this talk, I will outline possible ways to go about these challenges to tackle scarce data and label bias. I will draw upon recent research in cross-lingual learning, data selection and learning from disagreement and present (on-going) work applied to NLP tasks such as syntactic processing, named entity recognition and task-oriented dialogue, showing how weak supervision and multi-task learning can help remedy some of these challenges.

**Ben Hutchinson: Putting NLP Ethics Into Context**

In order to consider the societal and ethical consequences of biases in NLP models, it is necessary to consider how the models will be integrated into user-facing AI systems and products. We also need to consider who those systems will be used by, on and with. In the first part of this talk, I will adopt a wide lens and consider technology ethics within various social, cultural and historical contexts, using examples from my research. In the second part of this talk, I will zoom in to discuss practical challenges that arise when building NLP systems that are contextually appropriate and responsible.

**Dirk Hovy: More than words – Integrating social factors into language modeling**

Language is a social construct. We use it to achieve various conversational goals. Only one among them is to convey information. However, natural language processing has traditionally focused only on this informational aspect, ignoring all social aspects of language. That restriction was partially necessary to make modeling progress. However, I argue that as modeling power increases, we might want to revisit the issue. Social aspects of language can help disambiguate meaning, add more nuance to our models, and are becoming increasingly important in all aspects of generation. In this talk, I will outline several of the social dimensions that influence language use, how they affect NLP models, and what efforts are already underway to incorporate them. I will conclude with some open questions and ideas for future directions. If we manage to include social aspects of language into NLP, I believe we will open new research avenues, improve performance, and create fairer language technology.

**8th December (Wednesday) Workshop Day 1 (all times AEDT)**

14:00 Tutorial on Machine Translation and Summarization
Inigo Jauregi, Jacob Parnell, and Massimo Piccardi

16:30 Social Gathering/Mentoring – Kumospace

**9th December (Thursday) Day 2**

9:00 ALTA Keynote – Barbara Plank: Tackling scarce and biased data for more inclusive Natural Language

10:00 Social Gathering/Brunch/Mentorship – Kumospace

10:30 Welcome to Day 2

10:40 Oral Presentations 1 (Long: 10m + 3m QA, Short: 8m + 2m)

*Combining Shallow and Deep Representations for Text-Pair Classification* (long)
Vincent Nguyen, Sarvnaz Karimi and Zhenchang Xing

*Robustness Analysis of Grover for Machine-Generated News Detection* (long)
Rinaldo Gagiano, Maria Myung-Hee Kim, Xiuzhen Zhang and Jennifer Biggs

*Using Word Embeddings to Quantify Ethnic Stereotypes in 12 years of Spanish News.* (long)
Danielly Sorato, Diana Zavala-Rojas and Maria del Carme Colominas Ventura

*Multi-modal Intent Classification for Assistive Robots with Large-scale Naturalistic Datasets* (long)
Karun Varghese Mathew, Venkata S Aditya Tarigoppula and Lea Frermann

*Does QA-based intermediate training help fine-tuning language models for text classification?* (short)
Shiwei Zhang and Xiuzhen Zhang

*Using Discourse Structure to Differentiate Entities in Literature* (short)
Antonio Jimeno Yepes, Ameer Albahem and Karin Verspoor

11:52 Afternoon Break/Lunch/Social Gathering/Mentorship

14:00 Oral Presentations 2 (Long: 10m + 3m QA, Short: 8m + 2m)

*An Approach to the Frugal Use of Human Annotators for Text Classification Tasks* (long)
Li'An Chen and Hanna Suominen

*Exploring the Vulnerability of NLP Models via Universal Adversarial Texts* (long)
Xinzhe Li, Ming Liu, Xingjun Ma and Longxiang Gao

*Phone Based Keyword Spotting for Transcribing Very Low Resource Languages* (long)
Eric Le Ferrand, Steven Bird and Laurent Besacier

*Exploring Story Generation with Multi-task Objectives in Variational Autoencoders* (long)
Zhuohan Xie, Jey Han Lau and Trevor Cohn

14:52 Break: Social Gathering/Mentorship

15:07 Oral Presentations 3 (Long: 10m + 3m QA, Short: 8m + 2m)

*Evaluating Hierarchical Document Categorisation* (short)
Qian Sun, A. Shen, H. Yoshikawa, C. Ma, D. Beck, T. Iwakura and T. Baldwin

*Cross-Domain Language Modeling: An Empirical Investigation* (short)
Vincent Nguyen, Sarvnaz Karimi, Maciej Rybinski and Zhenchang Xing

*BERT's The Word : Sarcasm Target Detection using BERT* (short)
Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra and David Eyers

*A Computational Acquisition Model for Multi-modal Word Learning from Scratch* (abstract)
Uri Berger, Gabriel Stanovsky, Omri Abend and Lea Frermann
*Retrodiction as Delayed Recurrence: the Case of Adjectives in Italian and English* (short)
Raquel G. Alhama, Francesca Zermiani and Atiqah Khaliq
*Automatic Post-Editing for Translating Chinese Novels to Vietnamese* (short)
Thanh Vu and Dai Quoc Nguyen

16:07 Poster Session (all papers in Oral presentations 1 and 2)
17:30 End of day 2

**10th December (Friday) Day 3**

10:00 Welcome to Day 3
10:05 Oral Presentations 4 (Long: 10m + 3m QA, Short: 8m + 2m)
*Inductive Biases for Low Data VQA: A Data Augmentation Approach* (abstract)
Narjes Askarian, Ehsan Abbasnejad, Ingrid Zukerman, Wray Buntine and Reza Haffari
*Evaluation of Review Summaries via Question-Answering* (long)
Nannan Huang and Xiuzhen Zhang
*Curriculum Learning Effectively Improves Low Data VQA* (long)
Narjes Askarian, Ehsan Abbasnejad, Ingrid Zukerman, Wray Buntine and Reza Haffari
*Document Level Hierarchical Transformer* (long)
Najam Zaidi, Trevor Cohn and Gholamreza Haffari

11:00 ALTA Keynote – Barbara Plank: Tackling scarce and biased data for more inclusive Natural Language
12:00 Break/Lunch/Social Gathering/Mentorship – Kumospace
13:30 Oral Presentations 5 (Long: 10m + 3m QA, Short: 8m + 2m)
*Harnessing Privileged Information for Hyperbole Detection* (long)
Rhys Biddle, Maciek Rybinski, Qian Li, Cecile Paris and Guandong Xu
*Principled Analysis of Energy Discourse with Topic Labeling* (long)
Thomas Scelsi, Alfonso Martinez Arranz and Lea Frermann
*Generating and Modifying Natural Language Explanations* (long)
Abdus Salam, Rolf Schwitter and Mehmet Orgun
*Findings on Conversation Disentanglement* (long)
Rongxin Zhu, Jey Han Lau and Jianzhong Qi

14:22 Oral Presentations 6 (shared task papers: 8m + 2m)
*Overview of the 2021 ALTA Shared Task: Automatic Grading of Evidence, 10 Years Later* (long)
Diego Molla-Aliod
*Quick, get me a Dr. BERT: Automatic Grading of Evidence using Transfer Learning* (long)
Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra, David Eyers
*An Ensemble Model for Automatic Grading of Evidence* (long)
Yuting Guo, Yao Ge, Ruqi Liao, Abeed Sarker
*Handling Variance of Pretrained Language Models in Grading Evidence in the Medical Literature* (long)
Fajri Koto, Biaoyan Fang

15:02 ALTA General Meeting
15:17 Best Paper Awards
15:22 Poster Session (All papers in Oral sessions 3 and 4)
17:00 ALTA Keynote – Dirk Hovy: More than words – Integrating social factors into language modeling
18:00 Social Gathering/Mentorship – Kumospace
16:30 End of Day 3

# Table of Contents

## Long Papers


Findings on Conversation Disentanglement . . . . . 1
*Rongxin Zhu, Jey Han Lau and Jianzhong Qi*

An Approach to the Frugal Use of Human Annotators to Scale up Auto-coding for Text Classification Tasks . 12
*Li'An Chen and Hanna Suominen*

Curriculum Learning Effectively Improves Low Data VQA . . . . . 22
*Narjes Askarian, Ehsan Abbasnejad, Ingrid Zukerman, Wray Buntine and Gholamreza Haffari*

Using Word Embeddings to Quantify Ethnic Stereotypes in 12 years of Spanish News . . . . . 34
*Danielly Sorato, Diana Zavala-Rojas and Maria del Carme Colominas Ventura*

Multi-modal Intent Classification for Assistive Robots with Large-scale Naturalistic Datasets . . . . . 47
*Karun Varghese Mathew, Venkata S Aditya Tarigoppula and Lea Frermann*

Harnessing Privileged Information for Hyperbole Detection . . . . . 58
*Rhys Biddle, Maciek Rybinski, Qian Li, Cecile Paris and Guandong Xu*

Combining Shallow and Deep Representations for Text-Pair Classification . . . . . 68
*Vincent Nguyen, Sarvnaz Karimi and Zhenchang Xing*

Phone Based Keyword Spotting for Transcribing Very Low Resource Languages . . . . . 79
*Eric Le Ferrand, Steven Bird and Laurent Besacier*

Evaluation of Review Summaries via Question-Answering . . . . . 87
*Nannan Huang and Xiuzhen Zhang*

Exploring Story Generation with Multi-task Objectives in Variational Autoencoders . . . . . 97
*Zhuohan Xie, Jey Han Lau and Trevor Cohn*

Principled Analysis of Energy Discourse across Domains with Thesaurus-based Automatic Topic Labeling . . 107
*Thomas Scelsi, Alfonso Martinez Arranz and Lea Frermann*

Robustness Analysis of Grover for Machine-Generated News Detection . . . . . 119
*Rinaldo Gagiano, Maria Myung-Hee Kim, Xiuzhen Zhang and Jennifer Biggs*

Document Level Hierarchical Transformer . . . . . 128
*Najam Zaidi, Trevor Cohn and Gholamreza Haffari*

Exploring the Vulnerability of Natural Language Processing Models via Universal Adversarial Texts . . . . . 138
*Xinzhe Li, Ming Liu, Xingjun Ma and Longxiang Gao*

Generating and Modifying Natural Language Explanations . . . . . 149
*Abdus Salam, Rolf Schwitter and Mehmet Orgun*


## Short Papers


Does QA-based intermediate training help fine-tuning language models for text classification? . . . . . 158
*Shiwei Zhang and Xiuzhen Zhang*

Retrodiction as Delayed Recurrence: the Case of Adjectives in Italian and English . . . . . 163
*Raquel G. Alhama, Francesca Zermiani and Atiqah Khaliq*

Automatic Post-Editing for Vietnamese . . . . . 169
*Thanh Vu and Dai Quoc Nguyen*

Using Discourse Structure to Differentiate Focus Entities from Background Entities in Scientific Literature . . 174
*Antonio Jimeno Yepes, Ameer Albahem and Karin Verspoor*

Evaluating Hierarchical Document Categorisation . . . . . 179
*Qian Sun, Aili Shen, Hiyori Yoshikawa, Chunpeng Ma, Daniel Beck, Tomoya Iwakura and Timothy Baldwin*

BERT's The Word : Sarcasm Target Detection using BERT . . . . . 185
*Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra and David Eyers*

Cross-Domain Language Modeling: An Empirical Investigation . . . . . 192
*Vincent Nguyen, Sarvnaz Karimi, Maciej Rybinski and Zhenchang Xing*

**Shared Task Papers**

Overview of the 2021 ALTA Shared Task: Automatic Grading of Evidence, 10 years later . . . . . 201
*Diego Mollá*

Quick, get me a Dr. BERT: Automatic Grading of Evidence using Transfer Learning . . . . . 205
*Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra and David Eyers*

An Ensemble Model for Automatic Grading of Evidence . . . . . 213
*Yuting Guo, Yao Ge, Ruqi Liao and Abeed Sarker*

Handling Variance of Pretrained Language Models in Grading Evidence in the Medical Literature . . . . . 218
*Fajri Koto and Biaoyan Fang*