EACL 2021

Innovative Use of NLP for Building Educational Applications

Proceedings of the 16th Workshop

April 20, 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 978-1-954085-11-4

Introduction

Welcome to the 16th year of the *Workshop on Innovative Use of NLP for Building Educational Applications* held in conjunction with the 16th Conference of the European Chapter of the Association of Computational Linguistics (EACL 2021). We are fortunate to be able to keep this tradition going – though virtually – with another year of inspiring research! Our 23 papers this year – to be presented as papers and posters – include topics related to education, including automated writing and evaluation, argument and passage detection, grammatical error correction, text simplification, readability and data analytics for text complexity, and NLP and Speech for learning and assessment. The conference will also hold a panel, entitled: *"New Challenges for Educational Technology in the Time of the Pandemic"* intended to shed light on the effect of the pandemic with regard to how we think about educational technology development.

This year we received a total of 44 submissions (excluding one withdrawn submission). We accepted 6 papers as oral presentations and 17 as poster presentations, for an overall acceptance rate of 52 % and 14 % based on only the acceptance of oral presentations. Oral presentations typically represent more mature research. Each paper was reviewed by three members of the Program Committee who were believed to be most appropriate for each paper. We continue to have a strong policy to deal with conflicts of interest. First, we continue to make a concerted effort to resolve conflicts of interest - specifically, we do not assign papers to a reviewer if the paper has an author from their institution. Second, organizing committee members recuse themselves from discussions about papers if there is a conflict of interest.

Papers are accepted on the basis of several factors, including the relevance to a core educational problem space, the novelty of the approach or domain, and the strength of the research. The accepted papers were highly diverse – an indicator of the growing variety of foci in this field. We continue to believe that the workshop framework designed to introduce work in progress and new ideas is important and we hope that the breadth and variety of research accepted for this workshop is represented.

The BEA16 workshop has presentations that apply NLP to topics related to education, as follows: **Automated writing and evaluation**: *Essay Quality Signals as Weak Supervision for Source-based Essay Scoring* (Zhang & Litman), *Automated Classification of Written Proficiency Levels on the CEFR-Scale through Complexity Contours and RNNs* (Kerz et al); **Argument and passage detection**: *Parsing Argumentative Structure in English-as-Foreign-Language Essays* (Putra, Teufel & Tokunaga), *"Sharks are not the threat humans are": Argument Component Segmentation in School Student Essays* (Alhindi & Ghosh), *Training and Domain Adaptation for Supervised Text Segmentation* (Glavaš et al); **Grammatical error correction**: *Synthetic Data Generation for Grammatical Error Correction with Tagged Corruption Models* (Stahlberg & Kumar), *Document-level Grammatical Error Correction* (Yuan & Bryant), *Data Strategies for Low-Resource Grammatical Error Correction* (Flachs et al), and *Assessing Grammatical Correctness in Language Learning* (Katinskaia & Yangarber); **Text simplification, readability and data analytics for text complexity**: *Text Simplification by Tagging* (Omelianchuk et al), *Broad Linguistic Complexity Analysis for Greek Readability Classification* (Chatzipanagiotidis et al), *Towards a Data Analytics Pipeline for the Visualisation of Complexity Metrics in L2 Writings* (Gaillat et al); **NLP and Speech for learning and assessment**: *Employing Distributional Semantics to Organize Task-focused Vocabulary Learning* (Ponnusamy & Meurers), *Detecting Negative Language Transfer Using Models that Represent Language Structure* (Wanderley & Epp), *Teacher's Tools for Estonian as a Second Language in Language Portal* (Sõnaveeb et al), *On the Application of Transformers for Estimating the Difficulty of Multiple-Choice Questions from Text* (Benedetto et al), *Automatically Generating Cause-and-Effect Questions from Passages* (Stasaski et al), **Interventions Recommendation**: *Professionals' Observations Analysis in Special Needs Education* (Muñoz & Bravo-Marquez), *C-Test Collector: A Proficiency Testing Application to Collect Training Data for C-Tests* (Haring et al), *Virtual Pre-Training Teacher Assessment and Feedback via Conversational Agents* (Datta et al), *Using Linguistic Features to Predict the Response Process Complexity Associated with Answering Clinical MCQs* (Yaneva et

al), *Character Set Construction for Chinese Language Learning* (Yeung & Lee), and *Negation Scope Resolution for Chinese as a Second Language* (Zhang et al).

We wish to thank everyone who showed interest and submitted a paper, all of the authors for their contributions, the members of the Program Committee for their thoughtful reviews, and everyone who is attending this workshop, virtually!

Finally, our special thanks go to the emergency reviewers who stepped in to provide their expertise and help ensure the highest level of feedback: we acknowledge the help of Beata Beigman Klebanov and Haley Lepp.

Jill Burstein, Educational Testing Service
Andrea Horbach, University of Duisburg-Essen
Ekaterina Kochmar, University of Cambridge
Ronja Laarmann-Quante, University of Duisburg-Essen
Claudia Leacock, Grammarly
Nitin Madnani, Educational Testing Services
Ildikó Pilán, University of Oslo
Helen Yannakoudakis, King's College London
Torsten Zesch, University of Duisburg-Essen

Organizers:

Jill Burstein, Educational Testing Service
Andrea Horbach, University of Duisburg-Essen
Ekaterina Kochmar, University of Cambridge
Ronja Laarmann-Quante, University of Duisburg-Essen
Claudia Leacock, Grammarly
Nitin Madnani, Educational Testing Services
Ildikó Pilán, University of Oslo
Helen Yannakoudakis, King's College London
Torsten Zesch, University of Duisburg-Essen

Program Committee:

Tazin Afrin, University of Pittsburgh
David Alfter, University of Gothenburg
Jason Angel, Instituto Politécnico Nacional
Piper Armstrong, Brigham Young University
Timo Baumann, Universität Hamburg
Lee Becker, Pearson
Kay Berkling, DHBW Cooperative State University Karlsruhe
Chris Bryant, University of Cambridge
Guanliang Chen, Monash University
Mei-Hua Chen, Department of Foreign Languages and Literature
Leshem Choshen, Hebrew University of Jerusalem
Mark Core, University of Southern California
Scott Crossley, Georgia State University
Orphée De Clercq, LT3, Ghent University
Kordula De Kuthy, Tübingen University
Carrie Demmans Epp, University of Alberta
Ann Devitt, Trinity College, Dublin
Mariano Felice, University of Cambridge
Michael Flor, Educational Testing Service
Thomas François, Université Catholique de Louvain
Jennifer-Carmen Frey, Eurac Research
Michael Gamon, Microsoft Research
Cyril Goutte, National Research Council Canada
Masato Hagiwara, Octanove Labs
Jiangang Hao, Educational Testing Service
Marti Hearst, University of California, Berkeley
Trude Heift, Simon Fraser University
Heiko Holz, LEAD Graduate School & Research Network
Chung-Chi Huang, Frostburg State University
Yi-Ting Huang, Academia Sinica
Radu Tudor Ionescu, University of Bucharest
Elma Kerz, RWTH Aachen University
Fazel Keshtkar, St. John's University
Mamoru Komachi, Tokyo Metropolitan University
Ji-Ung Lee, UKP Lab, TU Darmstadt

Diane Litman, University of Pittsburgh
Zitao Liu, TAL Education Group
Peter Ljunglöf, University of Gothenburg; Chalmers University of Technology
Anastassia Loukina, Educational Testing Service
Fabiana MacMillan, Rosetta Stone
Lieve Macken, Ghent University
Montse Maritxalar, University of the Basque Country
James Martin, University of Colorado Boulder
Sandeep Mathias, IIT Bombay
Ditty Mathew, IIT Madras
Stephen Mayhew, Duolingo
Julie Medero, Harvey Mudd College
Detmar Meurers, University of Tübingen
Natawut Monaikul, University of Illinois at Chicago
Farah Nadeem, University of Washington
Diane Napolitano, Associated Press
Hwee Tou Ng, National University of Singapore
Huy Nguyen, LingoChamp
Ulrike Pado, HFT Stuttgart
Long Qin, Singsound Inc
Mengyang Qiu, University at Buffalo
Marti Quixal, University of Tübingen
Taraka Rama, University of North Texas
Lakshmi Ramachandran, A9 (Amazon)
Hanumant Redkar, IIT Bombay
Robert Reynolds, Brigham Young University
Brian Riordan, Educational Testing Service
Alla Rozovskaya, City University of New York
Helmer Strik, Radboud University Nijmegen
Jan Švec, University of West Bohemia
Anaïs Tack, UCLouvain & KU Leuven
Yuen-Hsien Tseng, National Taiwan University
Shalaka Vaidya, Research assistant
Sowmya Vajjala, National Research Council, Canada
Giulia Venturi, Institute for Computational Linguistics
Tatiana Vodolazova, University of Alicante
Elena Volodina, University of Gothenburg, Sweden
Shuting Wang, Facebook
Zarah Weiss, University of Tübingen
Michael White, The Ohio State University; Facebook AI
Alistair Willis, Open University, UK
Victoria Yaneva, NBME; University of Wolverhampton
Seid Muhie Yimam, University of Hamburg
Mingzhi Yu, University of Pittsburgh
Zheng Yuan, University of Cambridge
Klaus Zechner, Educational Testing Service
Fabian Zehner, DIPF, Leibniz Institute for Research and Information in Education

Table of Contents

Negation Scope Resolution for Chinese as a Second Language
Mengyu Zhang, Weiqi Wang, Shuqiao Sun and Weiwei Sun 1

Text Simplification by Tagging
Kostiantyn Omelianchuk, Vipul Raheja and Oleksandr Skurzhanskyi 11

Employing distributional semantics to organize task-focused vocabulary learning
Haemanth Santhi Ponnusamy and Detmar Meurers 26

Synthetic Data Generation for Grammatical Error Correction with Tagged Corruption Models
Felix Stahlberg and Shankar Kumar 37

Broad Linguistic Complexity Analysis for Greek Readability Classification
Savvas Chatzipanagiotidis, Maria Giagkou and Detmar Meurers 48

Character Set Construction for Chinese Language Learning
Chak Yan Yeung and John Lee 59

Identifying negative language transfer in learner errors using POS information
Leticia Farias Wanderley and Carrie Demmans Epp 64

Document-level grammatical error correction
Zheng Yuan and Christopher Bryant 75

Essay Quality Signals as Weak Supervision for Source-based Essay Scoring
Haoran Zhang and Diane Litman 85

Parsing Argumentative Structure in English-as-Foreign-Language Essays
Jan Wira Gotama Putra, Simone Teufel and Takenobu Tokunaga 97

Training and Domain Adaptation for Supervised Text Segmentation
Goran Glavaš, Ananya Ganesh and Swapna Somasundaran 110

Data Strategies for Low-Resource Grammatical Error Correction
Simon Flachs, Felix Stahlberg and Shankar Kumar 117

Towards a Data Analytics Pipeline for the Visualisation of Complexity Metrics in L2 writings
Thomas Gaillat, Anas Knefati and Antoine Lafontaine 123

Estonian as a Second Language Teacher's Tools
Tiiu Üksik, Jelena Kallas, Kristina Koppel, Katrin Tsepelina and Raili Pool 130

Assessing Grammatical Correctness in Language Learning
Anisia Katinskaia and Roman Yangarber 135

On the application of Transformers for estimating the difficulty of Multiple-Choice Questions from text
Luca Benedetto, Giovanni Aradelli, Paolo Cremonesi, Andrea Cappelli, Andrea Giussani and Roberto Turrin 147

Automatically Generating Cause-and-Effect Questions from Passages
Katherine Stasaski, Manav Rathod, Tony Tu, Yunfang Xiao and Marti A. Hearst 158

Interventions Recommendation: Professionals' Observations Analysis in Special Needs Education
Javier Muñoz and Felipe Bravo-Marquez 171

C-Test Collector: A Proficiency Testing Application to Collect Training Data for C-Tests
Christian Haring, Rene Lehmann, Andrea Horbach and Torsten Zesch 180

Virtual Pre-Service Teacher Assessment and Feedback via Conversational Agents
Debajyoti Datta, Maria Phillips, James P. Bywater, Jennifer Chiu, Ginger S. Watson, Laura Barnes and Donald Brown 185

Automated Classification of Written Proficiency Levels on the CEFR-Scale through Complexity Contours and RNNs
Elma Kerz, Daniel Wiechmann, Yu Qiao, Emma Tseng and Marcus Ströbel 199

"Sharks are not the threat humans are": Argument Component Segmentation in School Student Essays
Tariq Alhindi and Debanjan Ghosh 210

Using Linguistic Features to Predict the Response Process Complexity Associated with Answering Clinical MCQs
Victoria Yaneva, Daniel Jurich, Le An Ha and Peter Baldwin 223

Conference Program

10:10 - 10:20 Opening remarks
10:20 - 11:00 Session 1
11:00 - 11:30 Coffee Break
11:30 - 12:10 Session 2
12:10 - 13:30 Poster Session 1
13:30 - 15:00 Lunch Break
15:00 - 15:40 Session 3
15:40 - 16:50 Poster Session 2
16:50 - 17:50 Panel
17:50 - 18:00 Closing remarks