

MIA 2022

The Workshop on Multilingual Information Access (MIA)

Proceedings of the Workshop

July 15, 2022

©2022 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 978-1-955917-89-6

Organizing Committee

Organizing Committee

Akari Asai, University of Washington
Eunsol Choi, University of Texas at Austin
Jonathan H. Clark, Google Research
Junjie Hu, University of Wisconsin-Madison
Chia-Hsuan Lee, University of Washington
Jungo Kasai, University of Washington
Shayne Longpre, MIT
Ikuya Yamada, Studio Ousia and RIKEN
Rui Zhang, Penn State University

Program Committee

Senior Program Committee

C. M. Downey, University of Washington
Mehrad Moradshahi, Stanford University
Mihaela Bornea, IBM, International Business Machines
Mihir Kale, Carnegie Mellon University
Ryokan Ri, The University of Tokyo, Tokyo Institute of Technology
Zhucheng Tu, Apple
Akari Asai, Paul G. Allen School of Computer Science & Engineering, University of Washington
Akiko Eriguchi, Microsoft
Chia-Hsuan Lee, University of Washington, Seattle
He Bai, University of Waterloo
Jessica Ouyang, University of Texas at Dallas
Jonathan H. Louis, Google
Mikio Nakano, Honda Research Institute
Rebecca J. Passonneau, Penn State University
Gabriel Skantze, KTH
Manfred Stede, Universität Potsdam
David Traum, University of Southern California
Koichiro Yoshino, Nara Institute of Science and Technology

Program Committee

Sean Andrist, Microsoft Research, United States
Masahiro Araki, Kyoto Institute of Technology, Japan
Ron Artstein, USC Institute for Creative Technologies, United States
Yoav Artzi, Cornell University, United States
Timo Baumann, Universität Hamburg, Germany
Frederic Bechet, Aix Marseille Université - LIS/CNRS, France
Steve Beet, Aculab plc, United Kingdom
Jose Miguel Benedi, Universitat Politècnica de València, Spain
Luciana Benotti, Universidad Nacional de Córdoba, Argentina
Yonatan Bisk, Carnegie Mellon University, United States
Nate Blaylock, Cerence, United States
Dan Bohus, Microsoft Research, United States
Johan Boye, KTH, Sweden
Chloé Braud, IRIT - CNRS, France
Hendrik Buschmeier, Bielefeld University, Germany
Andrew Caines, University of Cambridge, United Kingdom
Christophe Cerisara, Université de Lorraine, CNRS, LORIA, France
Senthil Chandramohan, Microsoft, United States
Lin Chen, Head of AI, Cambia Health Solutions, United States
Paul Crook, Facebook, United States
Heriberto Cuayahuitl, University of Lincoln, United Kingdom
Nina Dethlefs, University of Hull, United Kingdom
David DeVault, University of Southern California, United States
Barbara Di Eugenio, University of Illinois at Chicago, United States
Jens Edlund, KTH Speech, Music and Hearing, Sweden

Maxine Eskenazi, Carnegie Mellon University, United States
Keelan Evanini, Educational Testing Service, United States
Mauro Falcone, Fondazione Ugo Bordoni, Italy
Michel Galley, Microsoft Research, United States
Milica Gasic, Heinrich Heine University Duesseldorf, Germany
Kallirroi Georgila, University of Southern California, ICT, United States
Alborz Geramifard, Facebook AI, United States
Debanjan Ghosh, Educational Testing Service, United States
Jonathan Ginzburg, Université Paris-Diderot (Paris 7), France
Joakim Gustafson, KTH, Sweden
Ivan Habernal, Technische Universität Darmstadt, Germany
Helen Hastie, Heriot-Watt University, United Kingdom
Michael Heck, Heinrich Heine University, Germany
Behnam Hedayatnia, Amazon, United States
Ryuichiro Higashinaka, NTT Media Intelligence Labs., Japan
Takuya Hiraoka, NEC Central Research Laboratories, Japan
Thomas Howard, University of Rochester, United States
David M. Howcroft, Heriot-Watt University, United Kingdom
Ruihong Huang, Texas A&M University, United States
Michimasa Inaba, The University of Electro-Communications, Japan
Koji Inoue, Kyoto University, Japan
Filip Jurcicek, Apple Inc., United Kingdom
Tatsuya Kawahara, Kyoto University, Japan
Chris Kedzie, Columbia University, United States
Simon Keizer, Toshiba Research Europe Ltd, United Kingdom
Chandra Khatri, Senior AI Research Scientist, Uber AI, United States
Alexander Koller, Saarland University, Germany
Kazunori Komatani, Osaka University, Japan
Ivana Kruijff-Korbayova, DFKI, Germany
Kornel Laskowski, Carnegie Mellon University, United States
Fabrice Lefevre, Avignon Univ., France
Oliver Lemon, Heriot-Watt University, United Kingdom
Junyi Jessy Li, University of Texas at Austin, United States
Pierre Lison, Norwegian Computing Centre, Norway
Bing Liu, Facebook, United States
Eduardo Lleida Solano, University of Zaragoza, Spain
Ramon Lopez-Cozar, University of Granada, Spain
Nurul Lubis, Heinrich Heine University, Germany
Ross Mead, Semio, United States
Teruhisa Misu, Honda Research Institute USA, United States
Seungwhan Moon, Facebook Conversational AI, United States
Raymond Mooney, University of Texas at Austin, United States
Elena Musi, University of Liverpool, United Kingdom
Satoshi Nakamura, Nara Institute of Science and Technology and RIKEN AIP Center, Japan
Vincent Ng, University of Texas at Dallas, United States
Douglas O'Shaughnessy, INRS-EMT (Univ. of Quebec), Canada
Alexandros Papangelis, Uber AI, United States
Cecile Paris, CSIRO, Australia
Nanyun Peng, University of Southern California, United States
Laura Perez-Beltrachini, School of Informatics, University of Edinburgh, United Kingdom
Paul Piwek, The Open University, United Kingdom

Heather Pon-Barry, Mount Holyoke College, United States
Andrei Popescu-Belis, HEIG-VD / HES-SO, Switzerland
Abhinav Rastogi, Google Research, United States
Ehud Reiter, University of Aberdeen, United Kingdom
Norbert Reithinger, DFKI GmbH, Germany
Antonio Roque, Tufts University, United States
Carolyn Rose, Carnegie Mellon University, United States
Clayton Rothwell, Infocitex Corp., United States
Sakriani Sakti, Nara Institute of Science and Technology (NAIST) / RIKEN AIP, Japan
Ruhi Sarikaya, Amazon, United States
David Schlangen, University of Potsdam, Germany
Ethan Selfridge, Interactions LLC, United States
Georg Stemmer, Intel Corp., Germany
Matthew Stone, Rutgers University, United States
Svetlana Stoyanchev, Toshiba Europe, United Kingdom
Kristina Striegnitz, Union College, United States
Pei-Hao Su, PolyAI, United Kingdom
Hiroaki Sugiyama, NTT Communication Science Labs., Japan
António Teixeira, DETI/IEETA, University of Aveiro, Portugal
Takenobu Tokunaga, Tokyo Institute of Technology, Japan
Bo-Hsiang Tseng, University of Cambridge, United Kingdom
Gokhan Tur, Amazon Alexa AI, United States
Stefan Ultes, Mercedes-Benz AG, Germany
David Vandyke, Apple, United Kingdom
Hsin-Min Wang, Academia Sinica, Taiwan
Yi-Chia Wang, Uber AI, United States
Nigel Ward, University of Texas at El Paso, United States
Jason D Clark, Apple, United States
Jungo Kasai, Paul G. Allen School of Computer Science & Engineering, University of Washington
Sebastian Ruder, Google
Shayne Longpre, Massachusetts Institute of Technology
Tom Sherborne, University of Edinburgh
Tom Kwiatkowski, Google
Wei Wang, Apple AI/ML
Xinyan Yu, Department of Computer Science, University of Washington
Xutan Peng, University of Sheffield

Table of Contents

<i>Geographical Distance Is The New Hyperparameter: A Case Study Of Finding The Optimal Pre-trained Language For English-isiZulu Machine Translation.</i>	
Muhammad Umair Nasir and Innocent Amos Mchechesi	1
<i>An Annotated Dataset and Automatic Approaches for Discourse Mode Identification in Low-resource Bengali Language</i>	
Salim Sazed	9
<i>Pivot Through English: Reliably Answering Multilingual Questions without Document Retrieval</i>	
Ivan Montero, Shayne Longpre, Ni Lao, Andrew Frank and Christopher DuBois	16
<i>Cross-Lingual QA as a Stepping Stone for Monolingual Open QA in Icelandic</i>	
Vésteinn Snæbjarnarson and Hafsteinn Einarsson	29
<i>Multilingual Event Linking to Wikidata</i>	
Adithya Pratapa, Rishubh Gupta and Teruko Mitamura	37
<i>Complex Word Identification in Vietnamese: Towards Vietnamese Text Simplification</i>	
Phuong Nguyen and David Kauchak	59
<i>Benchmarking Language-agnostic Intent Classification for Virtual Assistant Platforms</i>	
Gengyu Wang, Cheng Qian, Lin Pan, Haode Qi, Ladislav Kunc and Saloni Potdar	69
<i>ZusammenQA: Data Augmentation with Specialized Models for Cross-lingual Open-retrieval Question Answering System</i>	
Chia-Chien Hung, Tommaso Green, Robert Litschko, Tornike Tsereteli, Sotaro Takeshita, Marco Bombieri, Goran Glavaš and Simone Paolo Ponzetto	77
<i>Zero-shot cross-lingual open domain question answering</i>	
Sumit Agarwal, Suraj Tripathi, Teruko Mitamura and Carolyn Penstein Rose	91
<i>MIA 2022 Shared Task Submission: Leveraging Entity Representations, Dense-Sparse Hybrids, and Fusion-in-Decoder for Cross-Lingual Question Answering</i>	
Zhucheng Tu and Sarguna Janani Padmanabhan	100
<i>MIA 2022 Shared Task: Evaluating Cross-lingual Open-Retrieval Question Answering for 16 Diverse Languages</i>	
Akari Asai, Shayne Longpre, Jungo Kasai, Chia-Hsuan Lee, Rui Zhang, Junjie Hu, Ikuya Yamada, Jonathan H. Clark and Eunsol Choi	108

Program

Friday, July 15, 2022

- 09:15 - 09:00 *Opening Remark*
- 10:15 - 09:15 *Invited talks for the model track*
- 11:00 - 10:15 *Model panels*
- 12:00 - 11:00 *Poster session*
- 13:00 - 12:00 *Lunch break*
- 13:45 - 13:00 *Shared task session*
- 14:00 - 13:45 *Best paper talk*
- 15:00 - 14:00 *Invited talks for the resource track*
- 15:15 - 15:00 *Break*
- 16:00 - 15:15 *Resource panels*
- 16:15 - 16:00 *Closing session*