

Introduction to the 1st Workshop on Natural Language Processing for COVID-19 at ACL 2020

*Karin Verspoor, Kevin Bretonnel Cohen, Mark Dredze, Emilio Ferrara,
Jonathan May, Robert Munro, Cecile Paris, Byron Wallace*

The unprecedented global pandemic related to the spread of the coronavirus SARS-COV-2 and the associated outbreak of the infection dubbed COVID-19 has had dramatic impacts worldwide during 2020. Scientists around the globe have responded to the pandemic, hoping to make some contribution to understanding, tracking, modeling, and/or responding.

The ACL community can play a unique role in supporting research to combat COVID-19. Valuable insights and critical information may be contained in vast quantities of unstructured text and speech data. Thousands of previously published research articles (and those being published on a daily basis) on coronavirus may shape our understanding of the virus or support best practice clinical management of the disease. Analysis of millions of social media posts may help us understand how the public at large is responding to the outbreak. Identifying spreading misinformation can be critical to public health messaging. Automatic identification and organization of helpful information collected from the web might aid public response.

The impetus behind organizing this “emergency” workshop was to highlight the myriad ways in which Natural Language Processing (NLP) could be used to respond to the COVID-19 pandemic, and the ACL community rose to the challenge, supported by resources such as the COVID-19 dataset from the Allen Institute for AI which was used for a Kaggle challenge¹. We are pleased to have one of the first papers introducing this important data set amongst our accepted papers [1].

We announced the workshop on April 03, 2020 and immediately created an OpenReview site open for submissions, following an open rolling review process in which we would review as papers were submitted. We opted for single-blind reviewing, so that papers would be visible online from submission, and reviewing could proceed in an open manner. Public commentary was also enabled on the submissions, to allow for ongoing discussion of the submitted work. We received our first paper on April 09 [11], indicating just how ready the NLP community was to respond.

In all, we received 75 submissions; 50 of these arrived in the final two days before submission closed, 9 days ahead of the workshop date. With the rush of last-minute submissions, we were overwhelmed – both by the tremendous response of the community to the call, and the daunting prospect of running a rigorous review process with barely a week’s turnaround to the workshop. We made the difficult decision to defer the final 50 submissions to a different process; Part 2 of the workshop is therefore now in preparation for EMNLP2020.

Of the 25 submissions that were reviewed, 17 (68%) were selected for presentation. All of these papers are included in this Proceedings volume in some form; either as an abstract only, or as a short or long paper. The topics they address range from literature mining to social media analysis.

We invited authors of the 50 deferred papers to submit posters or videos for their work, and several took us up on the opportunity. These are linked from the workshop website at <https://www.nlpccovid19workshop.org/acl2020/posters>. We also set up virtual “poster sessions” via Zoom and announced these on social media.

These months of emergency workshop organization proved to be intense but very rewarding; we are greatly appreciative to all of the authors who submitted their work, the reviewers who helped us assess submissions in a timely manner, and the broader efforts of the ACL community to encourage and enable us to pull everything together. We are proud to showcase the tremendous collective work in this Proceedings volume.

¹<https://www.kaggle.com/allen-institute-for-ai/COVID-19-research-challenge>

Program Committee

Organizers

Karin Verspoor, University of Melbourne (Australia)
Kevin Bretonnel Cohen, University of Colorado Anschutz Medical Center (USA)
Mark Dredze, Johns Hopkins University (USA)
Emilio Ferrara, University of Southern California (USA)
Jonathan May, Information Sciences Institute (USA)
Robert Munro, Machine Learning Consulting (USA)
Cecile Paris, CSIRO (Australia)
Byron Wallace, Northeastern University (USA)

Reviewers

We are grateful for the huge efforts of our reviewers, who responded to the need for rapid reviews in the unfamiliar OpenReview system, as well as engaging in more interactive discussion through several rounds of author response. Quite a number of reviewers were recruited urgently after the submission site closed on June 30, and were asked to review with a turnaround of a few days. We couldn't have reviewed even the 25 submissions considered in the available time without the strong response to our (slightly desperate) requests for assistance. So a big, "Thank you!" goes out to the following reviewers:

1. Abeed Sarker, Emory University (USA)
2. Aditya Joshi, CSIRO (Australia)
3. Alexander Spangher, University of Southern California (USA)
4. Antonio Jimeno Yepes, IBM Research (Australia)
5. Benjamin E. Nye, Northeastern University (USA)
6. Berry de Bruijn, National Research Council (Canada)
7. Bevan Koopman, CSIRO (Australia)
8. Christian Boitet, IMAG (France)
9. Daniel Santel, Cincinnati Children's Medical Center (USA)
10. Danielle L. Mowery, University of Pennsylvania (USA)
11. David Lowell, Northeastern University (USA)
12. David Martinez Iraola, IBM Research (Australia)
13. Davy Weissenbacher, University of Pennsylvania (USA)
14. Diego Molla, Macquarie University (Australia)
15. Estrid He, University of Melbourne (Australia)
16. Hadi Amiri, University of Massachusetts (Lowell)
17. Haewoon Kwak, Qatar Computing Research Institute (Qatar)
18. Helen L. Johnson, University of Colorado (USA)
19. Jari Björne, University of Turku (Finland)
20. Jay DeYoung, Northeastern University (USA)
21. Jimmy Lin, University of Waterloo (Canada)
22. Jingbo Xia, Intramural Institute of Applied Mathematics Research, HZAU (China)
23. Jonathan May, Information Sciences Institute (USA)
24. Kevin Bretonnel Cohen, University of Colorado Anschutz Medical Center (USA)

25. Kirk Roberts, University of Texas Health Science Center (USA)
26. Manirupa Das, The Ohio State University (USA)
27. Maria Liakata, Warwick University (UK)
28. Matthias Gallé, Naver Labs Europe (France)
29. Mayla Rachel Boguslav, University of Colorado Anschutz Medical Center (USA)
30. Meladel Mistica, University of Melbourne (Australia)
31. Michael Conway, University of Utah (USA)
32. Oliver Baclic, Public Health Agency of Canada (Canada)
33. Olivier Bodenreider, National Institutes of Health (USA)
34. Orin Hargraves, University of Colorado (USA)
35. Peter T. Corbett, Royal Society of Chemistry
36. Pierre Zweigenbaum, LIMSI (France)
37. Qian Hu, MITRE (USA)
38. Rezarta Islamaj, National Library of Medicine (USA)
39. Robert Leaman, National Institutes of Health (USA)
40. Sarthak Jain, Northeastern University (USA)
41. Sarvnaz Karimi, CSIRO (Australia)
42. Silvio Amir, Johns Hopkins University (USA)
43. Simon Suster, University of Melbourne (Australia)
44. Srijan Kumar, Georgia Tech (USA)
45. Timothy A Miller, Harvard University (USA)
46. Tristan Naumann, Microsoft (USA)
47. Vlada Rozova, University of Melbourne (Australia)
48. William R. Hersh, Oregon Health and Sciences University (USA)
49. Zenan Zhai, University of Melbourne (Australia)

Presentation Program

The papers selected for presentation at the workshop are listed below, in thematic groups. The type of paper included in the Proceedings is indicated by (Long), (Short), or (Abs) for Abstracts.

Pre-recorded videos of some of the papers listed below are available linked from the workshop website².

Literature Analysis and Retrieval

- [1] (Long) *CORD-19: The COVID-19 Open Research Dataset*.
Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darrin Eidedarrine, Kathryn Funk, Yannis Katsis, Rodney Kinney, Yunyao Li, Ziyang Liu, William Merrill, Paul Mooney, Dewey Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shenzhihosh, Brandon Stilson, Alex Wade, Kuansan Wang, Nancy Xin Ru Wang, Chris Wilhelm, Boya Xie, Douglas Raymond, Daniel S Weld, Oren Etzioni, Sebastian Kohlmeier.
- [2] (Abs) *Rapidly Deploying a Neural Search Engine for the COVID-19 Open Research Dataset: Preliminary Thoughts and Lessons Learned*.
Edwin Zhang, Nikhil Gupta, Rodrigo Nogueira, Kyunghyun Cho, Jimmy Lin.
- [3] (Abs) *Document Classification for COVID-19 Literature*.
Bernal Jiménez Gutiérrez, Juncheng Zeng, Dongdong Zhang, Ping Zhang, Yu Su.
- [4] (Long) *Enabling Low-Resource Transfer Learning across COVID-19 Corpora by Combining Event-Extraction and Co-Training*.
Alexander Spangher, Nanyun Peng, Jonathan May, Emilio Ferrara.
- [5] (Abs) *Self-supervised context-aware COVID-19 document exploration through atlas grounding*.
Dusan Grujicic, Gorjan Radevski, Tinne Tuytelaars, Matthew B. Blaschko.
- [6] (Long) *CODA-19: Reliably Annotating Research Aspects on 10,000+ CORD-19 Abs Using a Non-Expert Crowd*.
Ting-Hao (Kenneth) Huang, Chieh-Yang Huang, Chien-Kuang Cornelia Ding, Yen-Chia Hsu, C. Lee Giles.
- [7] (Abs) *Information Retrieval and Extraction on COVID-19 Clinical Articles Using Graph Community Detection and Bio-BERT Embeddings*.
Debasmita Das, Yatin Katyal, Janu Verma, Rajesh Kumar Ranjan, Shashank Dubey, Aakash Deep Singh, Sourojit Bhaduri, Kushagra Agarwal.

COVID Question Answering

- [8] (Short) *What Are People Asking About COVID-19? A Question Classification Dataset*.
Jerry Wei, Chengyu Huang, Soroush Vosoughi, Jason Wei.
- [9] (Abs) *Jennifer for COVID-19: An NLP-Powered Chatbot Built for the People and by the People to Combat Misinformation*.
Yunyao Li, Tyrone Grandison, Patricia Silveyra, Ali Douraghy, Xinyu Guan, Thomas Kieselbach, Chengkai Li, Haiqi Zhang.

²<https://www.nlpCOVID19workshop.org/acl2020/schedule>

Clinical and Mental Health

- [10] (Short) *A Natural Language Processing System for National COVID-19 Surveillance in the US Department of Veterans Affairs.*
Alec B Chapman, Kelly S Peterson, Augie Turano, Tamára L Box, Katherine S Wallace, Makoto Jones.
- [11] (Long) *Measuring Emotions in the COVID-19 Real World Worry Dataset.*
Bennett Kleinberg, Isabelle van der Vegt, Maximilian Mozes.
- [12] (Short) *Estimating the effect of COVID-19 on mental health: Linguistic indicators of depression during a global pandemic.*
JT Wolohan.

Social Media

- [13] (Short) *Exploration of Gender Differences in COVID-19 Discourse on Reddit.*
Jai Aggarwal, Ella Rabinovich, Suzanne Stevenson.
- [14] (Long) *Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic.*
Anna Kruspe, Matthias Haeberle, Xiao Xiang Zhu.
- [15] (Short) *Cross-lingual Transfer Learning for COVID-19 Outbreak Alignment.*
Sharon Levy, William Yang Wang.
- [16] (Long) *COVID-19 and Arabic Twitter: How can Arab World Governments and Public Health Organizations Learn from Social Media?*
Lama Alsudias, Paul Rayson.
- [17] (Short) *NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube.*
Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, Simon Hegelich.

References

- [1] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darrin Eide, Kathryn Funk, Yannis Katsis, Rodney Michael Kinney, Yunyao Li, Ziyang Liu, William Merrill, Paul Mooney, Dewey A. Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex D. Wade, Kuansan Wang, Nancy Xin Ru Wang, Christopher Wilhelm, Boya Xie, Douglas M. Raymond, Daniel S. Weld, Oren Etzioni, and Sebastian Kohlmeier. CORD-19: The COVID-19 open research dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [2] Edwin Zhang, Nikhil Gupta, Rodrigo Nogueira, Kyunghyun Cho, and Jimmy Lin. Rapidly deploying a neural search engine for the COVID-19 Open Research Dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [3] Bernal Jiménez Gutiérrez, Juncheng Zeng, Dongdong Zhang, Ping Zhang, and Yu Su. Document classification for COVID-19 literature. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [4] Alexander Spangher, Nanyun Peng, Jonathan May, and Emilio Ferrara. Enabling low-resource transfer learning across COVID-19 corpora by combining event-extraction and co-training. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [5] Dusan Grujicic, Gorjan Radevski, Tinne Tuytelaars, and Matthew Blaschko. Self-supervised context-aware COVID-19 document exploration through atlas grounding. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [6] Ting-Hao Kenneth Huang, Chieh-Yang Huang, Chien-Kuang Cornelia Ding, Yen-Chia Hsu, and C Lee Giles. CODA-19: Using a non-expert crowd to annotate research aspects on 10,000+ abstracts in the COVID-19 open research dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [7] Debasmita Das, Yatin Katyal, Janu Verma, Shashank Dubey, AakashDeep Singh, Kushagra Agarwal, Sourojit Bhaduri, and RajeshKumar Ranjan. Information retrieval and extraction on COVID-19 clinical articles using graph community detection and Bio-BERT embeddings. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [8] Jerry Wei, Chengyu Huang, Soroush Vosoughi, and Jason Wei. What are people asking about COVID-19? a question classification dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [9] Yunyao Li, Tyrone Grandison, Patricia Silveyra, Ali Douraghy, Xinyu Guan, Thomas Kieselbach, Chengkai Li, and Haiqi Zhang. Jennifer for covid-19: An NLP-powered chatbot built for the people and by the people to combat misinformation. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [10] Alec Chapman, Kelly Peterson, Augie Turano, Tamára Box, Katherine Wallace, and Makoto Jones. A natural language processing system for national COVID-19 surveillance in the us department of veterans affairs. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [11] Bennett Kleinberg, Isabelle van der Vegt, and Maximilian Mozes. Measuring Emotions in the COVID-19 Real World Worry Dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [12] JT Wolohan. Estimating the effect of COVID-19 on mental health: Linguistic indicators of depression during a global pandemic. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July. Association for Computational Linguistics.
- [13] Jai Aggarwal, Ella Rabinovich, and Suzanne Stevenson. Exploration of gender differences in COVID-19 discourse on reddit. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [14] Anna Kruspe, Matthias Häberle, Iona Kuhn, and Xiao Xiang Zhu. Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.

- [15] Sharon Levy and William Yang Wang. Cross-lingual transfer learning for COVID-19 outbreak alignment. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [16] Lama Alsudias and Paul Rayson. COVID-19 and arabic twitter: How can arab world governments and public health organizations learn from social media? In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.
- [17] Juan Carlos Medina Serrano and Orestis Papakyriakopoulos and Simon Hegelich. NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, volume 1, Online, July 2020. Association for Computational Linguistics.