BioNLP 2022 @ ACL 2022

Proceedings of the
21st Workshop on Biomedical Language Processing

May 26, 2022

©2022 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 978-1-955917-27-8

Reflectively looking into the future of biomedical language processing

Dina Demner-Fushman, Sophia Ananiadou, Kevin Bretonnel Cohen, Junichi Tsujii

The 2022 meeting of the Biomedical Natural Language Processing workshop at the Association for Computational Linguistics conference reminds us of the first such workshop at the 2002 conference. The twenty years that have passed since then have seen enormous growth in the BioNLP community, and now seems like a good time to take stock of where we have come over the course of those two decades.

Interest in scientific natural language processing started soon after the launch of the Sputnik satellite in 1957, when the Anglophone scientific world realized that there was quite a bit of good research being published in Russian that it hadn't been reading. Interest in, first, clinical and then more general biomedical language processing started in the 1960s, and biomedical language processing interest groups soon formed within the clinical, and later the bioinformatics, communities. The Association for Computational Linguistics BioNLP community came together in 2002 to answer the needs of a deeply interdisciplinary area of research focused on natural language processing and text mining methods applied to biomedical text. The field spread quickly, but the events and publication venues for computational linguists interested in the biomedical sublanguage were dispersed across a range of disciplines and conferences. Clinical natural language processing had a natural home in the Association for Medical Informatics, and biologically oriented language processing focused on the rapidly growing scientific literature was well-housed in the International Society for Molecular Biology and Pacific Symposium for Biocomputing publication venues; the Association for Computational Linguistics seemed like a natural home for research that focused around the linguistic nature of our field, rather than being oriented around its clinical and biological applications.

To bring together the passion for the domain and the benefits of belonging to the ACL community, SIGBioMed was formed as an ACL SIG in 2007. SIGBioMed is celebrating 15 years this summer. From the beginning, the SIG strove to be inclusive in terms of the topics of interest, languages studied, and researchers invited for presentations and keynotes. As can be seen in the work presented in this 2022 workshop, SIGBioMed continues that policy of diversity, equity and inclusion. Borrowing from the New York Times, SIGBioMed's (unofficial) motto is "All the Work That's Fit to Print—as long as it broadly applies to the biomedical and clinical domains."

Biomedical language processing started with rigorous text mining research that helped advancing understanding of biomedical text and provided services to the target domains. For example, the MedLee system was used to support clinical applications (Friedman et al., 2004), whereas BioNLP shared tasks in 2011 and 2013 focused on extraction of information about pathways and development of biomedical event extraction systems (Miwa, M. et al 2013; Björne J. et al. 2015). Recent developments in and availability of large pre-trained language models (BioBERT, ClinicalBERT, SciBERT, etc.) provide us not only with a chance to advance the research and applications towards language and context understanding, but also to start understanding how the models perform the tasks, as evidenced by the work presented in the next sections.

Looking back: The test of time award

This meeting marking two decades of research in and around the ACL community provides an opportunity to reflect on how we got to where we are. So, following up on a suggestion from Tim Miller, we solicited nominations for a new BioNLP Workshop award: recognition of papers in our field that have "stood the test of time."

As the nominations came in, we quickly realized that the request was underspecified. In what publication venues could a nominee have appeared? Should it be limited to the BioNLP Workshop, or would any venue qualify? How *much* time? Would authors be allowed to nominate their own papers? And what would it mean to have "stood the test," exactly? Number of citations? Actual usage of a system, a technique, a resource, an idea? Would a once-heavily-cited paper that is not cited much any more qualify? And did we need to normalize for the length of time since publication? And who should do the selection from amongst the nominees? What if a paper by a member of the organizing committee was nominated? The organizing committee? An external panel? Open vote of the entire community?

We began with the assumption that we might be able to induce the answers to those questions from the nominations themselves. To facilitate that, we asked nominators to consider writing up a note—of the length of their choice—describing why they felt that their nominee rated recognition. We explicitly allowed self-nomination. And then we waited.

As it turned out, the set of nominations did not answer our questions. Some nominators expressed a well-argued opinion that only publications from the BioNLP Workshop should qualify, but submissions came in from a number of venues. The time spans since their publications varied widely. They covered systems, techniques, resources, and—thank goodness—ideas. Papers were submitted by non-authors, papers were submitted by their own authors, and we had submissions that were co-authored by the organizers of the workshop.

The only thing we did *not* have was a voting mechanism. We considered counting the number of nominations per paper, but several papers were nominated twice; the only one that was nominated three times had two of its nominations from its own authors—not forbidden, but it made the number-of-nominations criterion seem unreasonable; and in any case, elementary power calculations soon convinced us that the total number of nominations was not sufficient to differentiate between one vote, two votes, or three. In the end, we contemplated the set of nominations, saw perfectly good reasons to accept that they had *all*, in one or more ways, "stood the test of time." Consequently, this year we are awarding the BioNLP Test of Time Award to multiple papers—in fact, to all of the papers that were nominated. You will find them listed in Table 1, which accords to all co-authors concerned the right to add "2022 BioNLP Test Of Time Award Recipient" to their CVs. Although we resolved essentially none of the issues that we had identified, this was a tremendously fun exercise, and we look forward to excellent suggestions from the community as to how to answer the questions that we raise above, as well as how to do this next year in a more principled way without quite so glaring an appearance of conflict of interest.

Looking forward: Overview of the work in this volume

BioNLP 2022 received 59 valid submissions, of which 11 were accepted as oral presentations and 32 as posters.

The scope and the depth of the work in this volume reflects the growing rigor and maturity of biomedical language processing. True to the historical inclusiveness of the workshop, the processed text includes scientific publications, clinical notes, and other forms of formal and informal communications, primarily in English, but also in Bangla (Sazzed et al.), Spanish-Catalan (Amin et al.), Spanish (Carrino et al.) and Romanian (Mitrofan et al.)

Advances in literature processing are reflected in the work that presents end-to-end document level relation extraction that leverages coreference resolution and entity extraction (Giorgi et al.); linking citing sentences in a publication to the cited sentences in referenced sources (Roy et al.); and extracting design and evidence from the descriptions of Clinical Trials (Witte et al.)

Aronson, Alan R. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc. AMIA Symposium, p. 17, 2001.
Björne, Jari, Juho Heimonen, Filip Ginter, Antti Airola, Tapio Pahikkala, and Tapio Salakoski. Extracting complex biological events with rich graph-based feature sets. Proc. BioNLP 2009 Workshop Companion Volume (Shared Task), pp. 10-18. 2009.
Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J. Biomed Inform. 2001 Oct;34(5):301-10.
Kim, J-D., Tomoko Ohta, Yuka Tateisi, and Jun'ichi Tsujii. GENIA corpus–a semantically annotated corpus for bio-textmining. Bioinformatics 19, no. suppl_1 (2003): i180-i182.
Leaman, Robert, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. Proc. Biomedical Natural Language Processing, pp. 117-125. 2010.
Morgan, Alexander A., Zhiyong Lu, Xinglong Wang, Aaron M. Cohen, Juliane Fluck, Patrick Ruch, Anna Divoli, Katrin Fundel, Robert Leaman, Jörg Hakenberg, Chengjie Sun, Heng-hui Liu, Rafael Torres, Michael Krauthammer, William W Lau, Hongfang Liu, Chun-Nan Hsu, Martijn Schuemie, K Bretonnel Cohen, and Lynette Hirschman. Overview of BioCreative II gene normalization. Genome Biology 9, no. 2 (2008): 1-19.
Oronoz M, Gojenola K, Pérez A, de Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. J. Biomed Inform. 2015 Aug;56:318-32.
Pestian, John P., Michael Sorter, Brian Connolly, Kevin Bretonnel Cohen, Cheryl McCullumsmith, Jeffry T. Gee, Louis-Philippe Morency, Stefan Scherer, Lesley Rohlfs, and STM Research Group. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide and Life-Threatening Behavior 47, no. 1 (2017): 112-121.
Sarker, Abeed, Rachel Ginn, Azadeh Nikfarjam, Karen O'Connor, Karen Smith, Swetha Jayaraman, Tejaswi Upadhaya, and Graciela Gonzalez. Utilizing social media data for pharmacovigilance: a review. J. Biomed. Inf. 54 (2015): 202-212.
Tsuruoka, Yoshimasa, Yuka Tateishi, Jin-Dong Kim, Tomoko Ohta, John McNaught, Sophia Ananiadou, and Jun'ichi Tsujii. Developing a robust part-of-speech tagger for biomedical text. Panhellenic Conference on Informatics, pp. 382-392. Springer, Berlin, Heidelberg, 2005.
Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, Clark C. Negation's not solved: generalizability versus optimizability in clinical natural language processing. PLoS One. 2014 Nov 13;9(11):e112774.

Table 1: The 2022 BioNLP Test Of Time Awardees, in alphabetical order.

The biomedical domain and particularly clinical language processing suffers from a dearth of resources. The community is clearly addressing the need for annotated data by creating new datasets, data augmentation, and exploring approaches to reducing the need for data. We see many efforts in zero-, few-shot training, data augmentation and distant supervision: for causal precedence among chemical interactions (Liang et al.); information extraction (events, named entity, and relation extraction) (Papanikolaou et al., Wang et al., Khandelwal et al., Iinuma et al., Trieu et al., Dhrangadhariya et al., Watanabe et al., Sarrouti et al., Phan et al., Kim et al.); term normalization (Zeng et al.), summarization (Soleimani et al), and cross-lingual transfer (Amin et al.).

The new datasets introduced at BioNLP 2022 include the Medical Video Question Answering Shared Task data (Guota et al.) ; biomedical named-entity annotated corpus for Bangla (Sazzed et al.); ICD coding (Huang et al.); and curation of antibiotic-resistant genes (Chandak et al.)

We are happy to see many efforts on model understanding and analysis. This volume includes work on explaining model decisions on health-related online materials (Boissonnet et al.); explanations of medical coding predictions (Wood-Doughty et al.); entity memorization and recall in pretrained large LMs with positional prompting (Abaho et al.); inter-annotator agreement and its relation to model performance (Richie et al.); and a self-supervised pre-training approach for understanding genetic information (Cahyawijaya et al.).

We notice increased interest in complex tasks of language generation, summarization and question answering. Language generation was studied both in general (Yuan et al.) and for the specific tasks of dialogue generation (Naseem et al., Ngai et al.) and radiology report generation (Yan et al., Tang et al.), Work on summarization includes extractive/abstractive summarization of documents of varying length (Bishop et al.), aspect-based scientific document summarization (Soleimani et al.) and summarization as an approach to calculate seizure frequencies and dates of last seizure (Xie et al.). Question answering was explored on its own (Pappas et al.) and as a tool for risk prediction (Liang et al.), event extraction (Wang et al.), and explaining quality assessment of online materials (Boissonnet et al.)

Clinical language processing shows stable interest in ICD coding (Michalopoulos et al., Falis et al., Wood-Doughty et al.), risk score prediction (Lianf et al.) and the impact of de-identification (Vakili et al.)

The **Medical Video Question Answering Shared Task** co-located with BioNLP 2022 is described in the overview (Gupta et al.) that includes 8 technical reports submitted by the participating teams, in addition to the two papers presented as posters in the workshop(Li et al., Kusa et al.)

Last, but most certainly not least in this era of rampant mental health concerns, approaches to supporting mental health were studied in the works on analysis of speech disfluencies towards automated dementia detection (Farzana et al.) and dialogue generation for psychotherapeutic counselling (Das et al.)

Acknowledging the community

As always, we are deeply grateful to the authors of the submitted papers and to the reviewers (listed elsewhere in this volume) who produced three thorough and thoughtful reviews for each paper in a fairly short review period.

The quality of submitted work continues growing and the Organizers are truly grateful to our amazing Program Committee that helped us determine which work is ready to be presented and which will benefit from additional experiments and analyses suggested by the reviewers.

Finally, we thank everyone who nominated papers for the Test of Time Award—especially for their

well-reasoned and insightful discussions of *why* they chose those papers.

As in years past, we are looking forward to a productive workshop, and we hope that new collaborations and research will evolve, continuing contributions of our community to public health and well-being.

References

Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17–21.

Björne, Jari, Juho Heimonen, Filip Ginter, Antti Airola, Tapio Pahikkala, and Tapio Salakoski. Extracting complex biological events with rich graph-based feature sets. Proc. BioNLP 2009 Workshop Companion Volume (Shared Task), pp. 10-18. 2009.

Björne J, Salakoski T. TEES 2.2: Biomedical Event Extraction for Diverse Corpora. BMC Bioinformatics. 2015;16 Suppl 16(Suppl 16):S4.

Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform. 2001 Oct;34(5):301-10. doi: 10.1006/jbin.2001.1029. PMID: 12123149.

Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc. 2004 Sep-Oct;11(5):392-402.

Kim, J-D., Tomoko Ohta, Yuka Tateisi, and Jun'ichi Tsujii. GENIA corpus–a semantically annotated corpus for bio-textmining. Bioinformatics 19, no. suppl_1 (2003): i180-i182.

Leaman, Robert, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. Proc. 2010 workshop on biomedical natural language processing, pp. 117-125. 2010.

Morgan, Alexander A., Zhiyong Lu, Xinglong Wang, Aaron M. Cohen, Juliane Fluck, Patrick Ruch, Anna Divoli, Katrin Fundel, Robert Leaman, Jörg Hakenberg, Chengjie Sun, Heng-hui Liu, Rafael Torres, Michael Krauthammer, William W Lau, Hongfang Liu, Chun-Nan Hsu, Martijn Schuemie, K Bretonnel Cohen, and Lynette Hirschman. Overview of BioCreative II gene normalization. Genome Biology 9, no. 2 (2008): 1-19.

Oronoz M, Gojenola K, Pérez A, de Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. J Biomed Inform. 2015 Aug;56:318-32. doi: 10.1016/j.jbi.2015.06.016. Epub 2015 Jun 30. PMID: 26141794.

Pestian, John P., Michael Sorter, Brian Connolly, Kevin Bretonnel Cohen, Cheryl McCullumsmith, Jeffry T. Gee, Louis-Philippe Morency, Stefan Scherer, Lesley Rohlfs, and STM Research Group. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide and Life-Threatening Behavior 47, no. 1 (2017): 112-121.

Pyysalo, S., Ohta, T., Rak, R. et al. Overview of the Cancer Genetics and Pathway Curation tasks of BioNLP Shared Task 2013. BMC Bioinformatics 16, S2 (2015).

Makoto Miwa, Sampo Pyysalo, Tomoko Ohta and Sophia Ananiadou (2013). Wide coverage biomedical event extraction using multiple partially overlapping corpora. BMC Bioinformatics, 14:175

Sarker, Abeed, Rachel Ginn, Azadeh Nikfarjam, Karen O'Connor, Karen Smith, Swetha Jayaraman,

Tejaswi Upadhaya, and Graciela Gonzalez. Utilizing social media data for pharmacovigilance: a review. Journal of Biomedical Informatics 54 (2015): 202-212.

Tsuruoka, Yoshimasa, Yuka Tateishi, Jin-Dong Kim, Tomoko Ohta, John McNaught, Sophia Ananiadou, and Jun'ichi Tsujii. Developing a robust part-of-speech tagger for biomedical text. In Panhellenic Conference on Informatics, pp. 382-392. Springer, Berlin, Heidelberg, 2005.

Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, Clark C. Negation's not solved: generalizability versus optimizability in clinical natural language processing. PLoS One. 2014 Nov 13;9(11):e112774. doi: 10.1371/journal.pone.0112774. PMID: 25393544; PMCID: PMC4231086.

Organizing Committee

Organizers:

Dina Demner-Fushman, US National Library of Medicine
Kevin Bretonnel Cohen, University of Colorado School of Medicine, USA
Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK
Junichi Tsujii, National Institute of Advanced Industrial Science and Technology, Japan

Program Committee:

Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK
Saadullah Amin, Saarland University, Germany
Emilia Apostolova, Anthem, Inc.
Eiji Aramaki, University of Tokyo, Japan
Timothy Baldwin, University of Melbourne, Australia
Spandana Balumuri, National Institute of Technology Karnataka, India
Steven Bethard, University of Arizona, USA
Robert Bossy, Inrae, Université Paris Saclay, France
Berry de Bruijn, National Research Council Canada
Leonardo Campillos Llanos, Universidad Autonoma de Madrid, Spain
Fenia Christopoulou, National Centre for Text Mining and University of Manchester, UK
Kevin Bretonnel Cohen, University of Colorado School of Medicine, USA
Brian Connolly, Ohio, USA
Mike Conway, University of Utah, USA
Manirupa Das, Amazon Search, Seattle, WA, USA
Surabhi Datta, The University of Texas Health Science Center at Houston, USA
Dina Demner-Fushman, US National Library of Medicine
Dmitriy Dligach, Loyola University Chicago, USA
Kathleen C. Fraser, National Research Council Canada
Travis Goodwin, US National Library of Medicine
Natalia Grabar, CNRS, France
Cyril Grouin, LIMSI - CNRS, France
Tudor Groza, The Garvan Institute of Medical Research, Australia
Deepak Gupta, US National Library of Medicine
Sam Henry, Christopher Newport University, USA
William Hogan, UCSD, USA
Kexin Huang, Stanford University, USA
Brian Hur, University of Melbourne, Australia
Richard Jackson, AstraZeneca
Antonio Jimeno Yepes, IBM, Melbourne Area, Australia
Sarvnaz Karimi, CSIRO, Australia
Nazmul Kazi, Montana State University, USA
Won Gyu Kim, US National Library of Medicine
Ari Klein, University of Pennsylvania, USA
Roman Klinger, University of Stuttgart, Germany
Andre Lamurias, University of Lisbon, Portugal
Majid Latifi, National College of Ireland
Alberto Lavelli, FBK-ICT, Italy
Robert Leaman, US National Library of Medicine
Lung-Hao Lee, National Central University, Taiwan
Ulf Leser, Humboldt-Universität zu Berlin, Germany

Diwakar Mahajan, IBM Thomas J. Watson Research Center, USA
Mark-Christoph Müller, Heidelberg Institute for Theoretical Studies, Germany
Claire Nedellec, INRA, France
Guenter Neumann, DFKI, Saarland, Germany
Aurelie Neveol, LIMSI - CNRS, France
Mariana Neves, German Federal Institute for Risk Assessment, Germany
Yifan Peng, US National Library of Medicine
Francisco J. Ribadas-Pena, University of Vigo, Spain
Anthony Rios, The University of Texas at San Antonio, USA
Angus Roberts, The University of Sheffield, UK
Kirk Roberts, The University of Texas Health Science Center at Houston, USA
Roland Roller, DFKI GmbH, Berlin, Germany
Mourad Sarrouti, Sumitovant Biopharma, Inc., USA
Mario Sänger, Humboldt-Universität zu Berlin, Germany
Diana Sousa, University of Lisbon, Portugal
Michael Spranger, Sony, Tokyo, Japan
Peng Su, University of Delaware, USA
Madhumita Sushil, University of California, San Francisco, USA
Karin Verspoor, The University of Melbourne, Australia
Roger Wattenhofer, ETH Zurich, Switzerland
Leon Weber, Humboldt Universität Berlin, Germany
Nathan M. White, James Cook University, Australia
Davy Weissenbacher, University of Pennsylvania, USA
W John Wilbur, US National Library of Medicine
Amelie Wührl, University of Stuttgart, Germany
Dongfang Xu, Harvard University, USA
Shweta Yadav, University of Illinois Chicago, USA
Jingqing Zhang, Imperial College London, UK
Ayah Zirikly, Johns Hopkins University, USA
Pierre Zweigenbaum, LIMSI - CNRS, France

Additional Reviewers:

Lifang Li, King's College London, UK
Ling Luo, National Library of Medicine, USA
Faith Wavinya Mutinda, University of Tokyo, Japan
Peter Belcak, ETH Zurich, Switzerland
Po-ting Lai, National Library of Medicine, USA
Beni Egressy, ETH Zurich, Switzerland
Zhao Meng, ETH Zurich, Switzerland

Table of Contents

Explainable Assessment of Healthcare Articles with QA
Alodie Boissonnet, Marzieh Saeidi, Vassilis Plachouras and Andreas Vlachos 1

A sequence-to-sequence approach for document-level relation extraction
John Giorgi, Gary Bader and Bo Wang 10

Position-based Prompting for Health Outcome Generation
Micheal Abaho, Danushka Bollegala, Paula Williamson and Susanna Dodd 26

How You Say It Matters: Measuring the Impact of Verbal Disfluency Tags on Automated Dementia Detection
Shahla Farzana, Ashwin Deshpande and Natalie Parde 37

Zero-Shot Aspect-Based Scientific Document Summarization using Self-Supervised Pre-training
Amir Soleimani, Vassilina Nikoulina, Benoit Favre and Salah Ait Mokhtar 49

Data Augmentation for Biomedical Factoid Question Answering
Dimitris Pappas, Prodromos Malakasiotis and Ion Androutsopoulos 63

Slot Filling for Biomedical Information Extraction
Yannis Papanikolaou, Marlene Staib, Justin Joshua Grace and Francine Bennett 82

Automatic Biomedical Term Clustering by Learning Fine-grained Term Representations
Sihang Zeng, Zheng Yuan and Sheng Yu 91

BioBART: Pretraining and Evaluation of A Biomedical Generative Language Model
Hongyi Yuan, Zheng Yuan, Ruyi Gan, Jiaxing Zhang, Yutao Xie and Sheng Yu 97

Incorporating Medical Knowledge to Transformer-based Language Models for Medical Dialogue Generation
Usman Naseem, Ajay Bandi, Shaina Raza, Junaid Rashid and Bharathi Raja Chakravarthi 110

Memory-aligned Knowledge Graph for Clinically Accurate Radiology Image Report Generation
Sixing Yan 116

Simple Semantic-based Data Augmentation for Named Entity Recognition in Biomedical Texts
Uyen Phan and Nhung Nguyen 123

Auxiliary Learning for Named Entity Recognition with Multiple Auxiliary Biomedical Training Data
Taiki Watanabe, Tomoya Ichikawa, Akihiro Tamura, Tomoya Iwakura, Chunpeng Ma and Tsuneo Kato 130

SNP2Vec: Scalable Self-Supervised Pre-Training for Genome-Wide Association Study
Samuel Cahyawijaya, Tiezheng Yu, Zihan Liu, Xiaopu ZHOU, Tze Wing Tiffany MAK, Yuk Yu Nancy IP and Pascale Fung 140

Biomedical NER using Novel Schema and Distant Supervision
Anshita Khandelwal, Alok Kar, Veera Raghavendra Chikka and Kamalakar Karlapalem 155

Improving Supervised Drug-Protein Relation Extraction with Distantly Supervised Models
Naoki Iinuma, Makoto Miwa and Yutaka Sasaki 161

Named Entity Recognition for Cancer Immunology Research Using Distant Supervision
Hai-Long Trieu, Makoto Miwa and Sophia Ananiadou 171

Intra-Template Entity Compatibility based Slot-Filling for Clinical Trial Information Extraction
Christian Witte and Philipp Cimiano 178

Pretrained Biomedical Language Models for Clinical NLP in Spanish
Casimiro Pio Carrino, Joan Llop, Marc Pàmies, Asier Gutiérrez-Fandiño, Jordi Armengol-Estapé, Joaquín Silveira-Ocampo, Alfonso Valencia, Aitor Gonzalez-Agirre and Marta Villegas 193

Few-Shot Cross-lingual Transfer for Coarse-grained De-identification of Code-Mixed Clinical Texts
Saadullah Amin, Noon Pokaratsiri Goldstein, Morgan Wixted, Alejandro Garcia-Rudolph, Catalina Martínez-Costa and Guenter Neumann 200

VPAI_Lab at MedVidQA 2022: A Two-Stage Cross-modal Fusion Method for Medical Instructional Video Classification
Bin Li, Yixuan Weng, Fei Xia, Bin Sun and Shutao Li 212

GenCompareSum: a hybrid unsupervised summarization method using salience
Jennifer Bishop, Qianqian Xie and Sophia Ananiadou 220

BioCite: A Deep Learning-based Citation Linkage Framework for Biomedical Research Articles
Sudipta Singha Roy and Robert E. Mercer 241

Low Resource Causal Event Detection from Biomedical Literature
Zhengzhong Liang, Enrique Noriega-Atala, Clayton Morrison and Mihai Surdeanu 252

Overview of the MedVidQA 2022 Shared Task on Medical Video Question-Answering
Deepak Gupta and Dina Demner-Fushman 264

Inter-annotator agreement is not the ceiling of machine learning performance: Evidence from a comprehensive set of simulations
Russell Richie, Sachin Grover and Fuchiang (Rich) Tsui 275

Conversational Bots for Psychotherapy: A Study of Generative Transformer Models Using Domain-specific Dialogues
Avisha Das, Salih Selek, Alia R. Warner, Xu Zuo, Yan Hu, Vipina Kuttichi Keloth, Jianfu Li, W. Jim Zheng and Hua Xu 285

BEEDS: Large-Scale Biomedical Event Extraction using Distant Supervision and Question Answering
Xing David Wang, Ulf Leser and Leon Weber 298

Data Augmentation for Rare Symptoms in Vaccine Side-Effect Detection
Bosung Kim and Ndapa Nakashole 310

Improving Romanian BioNER Using a Biologically Inspired System
Maria Mitrofan and Vasile Pais 316

BanglaBioMed: A Biomedical Named-Entity Annotated Corpus for Bangla (Bengali)
Salim Sazzed 323

ICDBigBird: A Contextual Embedding Model for ICD Code Classification
George Michalopoulos, Michal Malyska, Nicola Sahar, Alexander Wong and Helen Chen 330

Doctor XAvIer: Explainable Diagnosis on Physician-Patient Dialogues and XAI Evaluation
Hillary Ngai and Frank Rudzicz 337

DISTANT-CTO: A Zero Cost, Distantly Supervised Approach to Improve Low-Resource Entity Extraction Using Clinical Trials Literature
Anjani Dhrangadhariya and Henning Müller 345

EchoGen: Generating Conclusions from Echocardiogram Notes
Liyan Tang, Shravan Kooragayalu, Yanshan Wang, Ying Ding, Greg Durrett, Justin F. Rousseau and Yifan Peng 359

Quantifying Clinical Outcome Measures in Patients with Epilepsy Using the Electronic Health Record
Kevin Xie, Brian Litt, Dan Roth and Colin A. Ellis 369

Comparing Encoder-Only and Encoder-Decoder Transformers for Relation Extraction from Biomedical Texts: An Empirical Study on Ten Benchmark Datasets
Mourad Sarrouti, Carson Tao and Yoann Mamy Randriamihaja 376

Utility Preservation of Clinical Text After De-Identification
Thomas Vakili and Hercules Dalianis 383

Horses to Zebras: Ontology-Guided Data Augmentation and Synthesis for ICD-9 Coding
Matúš Falis, Hang Dong, Alexandra Birch and Beatrice Alex 389

Towards Automatic Curation of Antibiotic Resistance Genes via Statement Extraction from Scientific Papers: A Benchmark Dataset and Models
Sidhant Chandak, Liqing Zhang, Connor Brown and Lifu Huang 402

Model Distillation for Faithful Explanations of Medical Code Predictions
Zach Wood-Doughty, Isabel Cachola and Mark Dredze 412

Towards Generalizable Methods for Automating Risk Score Calculation
Jennifer J Liang, Eric Lehman, Ananya Iyengar, Diwakar Mahajan, Preethi Raghavan, Cindy Y. Chang and Peter Szolovits 426

DoSSIER at MedVidQA 2022: Text-based Approaches to Medical Video Answer Localization Problem
Wojciech Kusa, Georgios Peikos, Óscar Espitia, Allan Hanbury and Gabriella Pasi 432

Workshop Program

May 26, 2022

9:00–9:10 Opening Remarks

9:10–10:30 Session 1: Question Answering, Discourse Structure and Clinical Applications

9:10–9:30 *Explainable Assessment of Healthcare Articles with QA*
Alodie Boissonnet, Marzieh Saeidi, Vassilis Plachouras and Andreas Vlachos

9:30–9:50 *A sequence-to-sequence approach for document-level relation extraction*
John Giorgi, Gary Bader and Bo Wang

9:50–10:10 *Position-based Prompting for Health Outcome Generation*
Micheal Abaho, Danushka Bollegala, Paula Williamson and Susanna Dodd

10:10–10:30 *How You Say It Matters: Measuring the Impact of Verbal Disfluency Tags on Automated Dementia Detection*
Shahla Farzana, Ashwin Deshpande and Natalie Parde

10:30–11:00 Coffee Break

11:00–12:30 Poster Session 1

Zero-Shot Aspect-Based Scientific Document Summarization using Self-Supervised Pre-training
Amir Soleimani, Vassilina Nikoulina, Benoit Favre and Salah Ait Mokhtar

Data Augmentation for Biomedical Factoid Question Answering
Dimitris Pappas, Prodromos Malakasiotis and Ion Androutsopoulos

Slot Filling for Biomedical Information Extraction
Yannis Papanikolaou, Marlene Staib, Justin Joshua Grace and Francine Bennett

Automatic Biomedical Term Clustering by Learning Fine-grained Term Representations
Sihang Zeng, Zheng Yuan and Sheng Yu

May 26, 2022 (continued)

BioBART: Pretraining and Evaluation of A Biomedical Generative Language Model
Hongyi Yuan, Zheng Yuan, Ruyi Gan, Jiaxing Zhang, Yutao Xie and Sheng Yu

Incorporating Medical Knowledge to Transformer-based Language Models for Medical Dialogue Generation
Usman Naseem, Ajay Bandi, Shaina Raza, Junaid Rashid and Bharathi Raja Chakravarthi

Memory-aligned Knowledge Graph for Clinically Accurate Radiology Image Report Generation
Sixing Yan

Simple Semantic-based Data Augmentation for Named Entity Recognition in Biomedical Texts
Uyen Phan and Nhung Nguyen

Auxiliary Learning for Named Entity Recognition with Multiple Auxiliary Biomedical Training Data
Taiki Watanabe, Tomoya Ichikawa, Akihiro Tamura, Tomoya Iwakura, Chunpeng Ma and Tsuneo Kato

SNP2Vec: Scalable Self-Supervised Pre-Training for Genome-Wide Association Study
Samuel Cahyawijaya, Tiezheng Yu, Zihan Liu, Xiaopu ZHOU, Tze Wing Tiffany MAK, Yuk Yu Nancy IP and Pascale Fung

Biomedical NER using Novel Schema and Distant Supervision
Anshita Khandelwal, Alok Kar, Veera Raghavendra Chikka and Kamalakar Karlapalem

Improving Supervised Drug-Protein Relation Extraction with Distantly Supervised Models
Naoki Iinuma, Makoto Miwa and Yutaka Sasaki

Named Entity Recognition for Cancer Immunology Research Using Distant Supervision
Hai-Long Trieu, Makoto Miwa and Sophia Ananiadou

Intra-Template Entity Compatibility based Slot-Filling for Clinical Trial Information Extraction
Christian Witte and Philipp Cimiano

Pretrained Biomedical Language Models for Clinical NLP in Spanish
Casimiro Pio Carrino, Joan Llop, Marc Pàmies, Asier Gutiérrez-Fandiño, Jordi Armengol-Estapé, Joaquín Silveira-Ocampo, Alfonso Valencia, Aitor Gonzalez-Agirre and Marta Villegas

Few-Shot Cross-lingual Transfer for Coarse-grained De-identification of Code-Mixed Clinical Texts
Saadullah Amin, Noon Pokaratsiri Goldstein, Morgan Wixted, Alejandro Garcia-Rudolph, Catalina Martínez-Costa and Guenter Neumann

May 26, 2022 (continued)

VPAI_Lab at MedVidQA 2022: A Two-Stage Cross-modal Fusion Method for Medical Instructional Video Classification
Bin Li, Yixuan Weng, Fei Xia, Bin Sun and Shutao Li

12:30–14:00 Lunch Break

14:00–15:00 Session 2: Summarization and text mining

GenCompareSum: a hybrid unsupervised summarization method using salience
Jennifer Bishop, Qianqian Xie and Sophia Ananiadou

BioCite: A Deep Learning-based Citation Linkage Framework for Biomedical Research Articles
Sudipta Singha Roy and Robert E. Mercer

Low Resource Causal Event Detection from Biomedical Literature
Zhengzhong Liang, Enrique Noriega-Atala, Clayton Morrison and Mihai Surdeanu

15:00–15:30 Coffee Break

15:30–17:00 Poster Session 2

Overview of the MedVidQA 2022 Shared Task on Medical Video Question-Answering
Deepak Gupta and Dina Demner-Fushman

Inter-annotator agreement is not the ceiling of machine learning performance: Evidence from a comprehensive set of simulations
Russell Richie, Sachin Grover and Fuchiang (Rich) Tsui

Conversational Bots for Psychotherapy: A Study of Generative Transformer Models Using Domain-specific Dialogues
Avisha Das, Salih Selek, Alia R. Warner, Xu Zuo, Yan Hu, Vipina Kuttichi Keloth, Jianfu Li, W. Jim Zheng and Hua Xu

BEEDS: Large-Scale Biomedical Event Extraction using Distant Supervision and Question Answering
Xing David Wang, Ulf Leser and Leon Weber

May 26, 2022 (continued)

Data Augmentation for Rare Symptoms in Vaccine Side-Effect Detection
Bosung Kim and Ndapa Nakashole

Improving Romanian BioNER Using a Biologically Inspired System
Maria Mitrofan and Vasile Pais

BanglaBioMed: A Biomedical Named-Entity Annotated Corpus for Bangla (Bengali)
Salim Sazzed

ICDBigBird: A Contextual Embedding Model for ICD Code Classification
George Michalopoulos, Michal Malyska, Nicola Sahar, Alexander Wong and Helen Chen

Doctor XAvIer: Explainable Diagnosis on Physician-Patient Dialogues and XAI Evaluation
Hillary Ngai and Frank Rudzicz

DISTANT-CTO: A Zero Cost, Distantly Supervised Approach to Improve Low-Resource Entity Extraction Using Clinical Trials Literature
Anjani Dhrangadhariya and Henning Müller

EchoGen: Generating Conclusions from Echocardiogram Notes
Liyan Tang, Shravan Kooragayalu, Yanshan Wang, Ying Ding, Greg Durrett, Justin F. Rousseau and Yifan Peng

Quantifying Clinical Outcome Measures in Patients with Epilepsy Using the Electronic Health Record
Kevin Xie, Brian Litt, Dan Roth and Colin A. Ellis

Comparing Encoder-Only and Encoder-Decoder Transformers for Relation Extraction from Biomedical Texts: An Empirical Study on Ten Benchmark Datasets
Mourad Sarrouti, Carson Tao and Yoann Mamy Randriamihaja

Utility Preservation of Clinical Text After De-Identification
Thomas Vakili and Hercules Dalianis

Horses to Zebras: Ontology-Guided Data Augmentation and Synthesis for ICD-9 Coding
Matúš Falis, Hang Dong, Alexandra Birch and Beatrice Alex

Towards Automatic Curation of Antibiotic Resistance Genes via Statement Extraction from Scientific Papers: A Benchmark Dataset and Models
Sidhant Chandak, Liqing Zhang, Connor Brown and Lifu Huang

May 26, 2022 (continued)

Model Distillation for Faithful Explanations of Medical Code Predictions
Zach Wood-Doughty, Isabel Cachola and Mark Dredze

Towards Generalizable Methods for Automating Risk Score Calculation
Jennifer J Liang, Eric Lehman, Ananya Iyengar, Diwakar Mahajan, Preethi Raghavan, Cindy Y. Chang and Peter Szolovits

DoSSIER at MedVidQA 2022: Text-based Approaches to Medical Video Answer Localization Problem
Wojciech Kusa, Georgios Peikos, Óscar Espitia, Allan Hanbury and Gabriella Pasi

17:00–17:10 Closing Remarks