

NAACL 2021

Second Workshop on Scholarly Document Processing

Proceedings of the Workshop

June 10, 2021

Online

©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 978-1-954085-28-2

Message from the Program Chairs

Welcome to the Second Workshop on Scholarly Document Processing (SDP) at NAACL 2021.

Next to keeping up with the growing literature in their own and related fields, scholars increasingly also need to rebut pseudo-science and disinformation. To address this challenge, computational work on enhancing search, summarization, and analysis of scholarly documents has flourished. However, the various strands of research on scholarly document processing remain fragmented. To reach to the broader NLP and AI/ML community, pool distributed efforts and enable shared access to published research, we held the 2nd Workshop on Scholarly Document Processing at NAACL 2021. The SDP workshop consisted of a research track and three Shared Tasks, geared towards easier access to scientific methods and results. <https://sdproc.org/2021/>

Organizing Committee

Iz Beltagy, Allen Institute for Artificial Intelligence, USA
Arman Cohan, Allen Institute for Artificial Intelligence, USA
Guy Feigenblat, IBM Research AI, Haifa Research Lab, Israel
Dayne Freitag, SRI International, San Diego, USA
Tirthankar Ghosal, Institute of Formal and Applied Linguistics, Charles University, Czech Republic
Keith Hall, Google AI, USA
Drahomira Herrmannova, Oak Ridge National Laboratory, USA
Petr Knoth, Open University, UK
Kyle Lo, Allen Institute for Artificial Intelligence, USA
Philipp Mayr, GESIS – Leibniz Institute for the Social Sciences, Germany
Robert Patton, Oak Ridge National Laboratory, USA
Michal Shmueli-Scheuer, IBM Research AI, Haifa Research Lab, Israel
Anita de Waard, Elsevier, USA
Kuansan Wang, Microsoft Research, USA
Lucy Lu Wang, Allen Institute for Artificial Intelligence, USA

Table of Contents

<i>Determining the Credibility of Science Communication</i> Isabelle Augenstein	1
<i>Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation</i> Soyeong Jeong, Jinheon Baek, ChaeHun Park and Jong Park	7
<i>Task Definition and Integration For Scientific-Document Writing Support</i> Hiromi Narimatsu, Kohei Koyama, Kohji Dohsaka, Ryuichiro Higashinaka, Yasuhiro Minami and Hirotoishi Taira	18
<i>Detecting Anatomical and Functional Connectivity Relations in Biomedical Literature via Language Representation Models</i> Ibrahim Burak Ozyurt, Joseph Menke, Anita Bandrowski and Maryann Martone	27
<i>The Biomaterials Annotator: a system for ontology-based concept annotation of biomaterials text</i> Javier Corvi, Carla Fuenteslópez, José Fernández, Josep Gelpi, Maria-Pau Ginebra, Salvador Capella-Guitierrez and Osnat Hakimi	36
<i>Keyphrase Extraction from Scientific Articles via Extractive Summarization</i> Chrysovalantis Giorgos Kontoulis, Eirini Papagiannopoulou and Grigorios Tsooumakas	49
<i>Argument Mining for Scholarly Document Processing: Taking Stock and Looking Ahead</i> Khalid Al Khatib, Tirthankar Ghosal, Yufang Hou, Anita de Waard and Dayne Freitag	56
<i>Bootstrapping Multilingual Metadata Extraction: A Showcase in Cyrillic</i> Johan Krause, Igor Shapiro, Tarek Saier and Michael Färber	66
<i>The Effect of Pretraining on Extractive Summarization for Scientific Documents</i> Yash Gupta, Pawan Sasanka Ammanamanchi, Shikha Bordia, Arjun Manoharan, Deepak Mittal, Ramakanth Pasunuru, Manish Shrivastava, Maneesh Singh, Mohit Bansal and Preethi Jyothi	73
<i>Finding Pragmatic Differences Between Disciplines</i> Lee Kezar and Jay Pujara	83
<i>Extractive Research Slide Generation Using Windowed Labeling Ranking</i> Athar Sefid, Prasenjit Mitra, Jian Wu and C Lee Giles	91
<i>LongSumm 2021: Session based automatic summarization model for scientific document</i> senci ying, Zheng Yan Zhao and wuhe zou	97
<i>CNLP-NITS @ LongSumm 2021: TextRank Variant for Generating Long Summaries</i> Darsh Kaushik, Abdullah Faiz Ur Rahman Khilji, Utkarsh Sinha and Partha Pakray	103
<i>Unsupervised document summarization using pre-trained sentence embeddings and graph centrality</i> Juan Ramirez-Orta and Evangelos Milios	110
<i>QMUL-SDS at SCIVER: Step-by-Step Binary Classification for Scientific Claim Verification</i> Xia Zeng and Arkaitz Zubiaga	116

Conference Program

Keynote Paper

Determining the Credibility of Science Communication

Isabelle Augenstein

Research Track: Long Papers

Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation

Soyeong Jeong, Jinheon Baek, ChaeHun Park and Jong Park

Task Definition and Integration For Scientific-Document Writing Support

Hiromi Narimatsu, Kohei Koyama, Kohji Dohsaka, Ryuichiro Higashinaka, Yasuhiro Minami and Hirotoishi Taira

Detecting Anatomical and Functional Connectivity Relations in Biomedical Literature via Language Representation Models

Ibrahim Burak Ozyurt, Joseph Menke, Anita Bandrowski and Maryann Martone

The Biomaterials Annotator: a system for ontology-based concept annotation of biomaterials text

Javier Corvi, Carla FuentesLópez, José Fernández, Josep Gelpi, Maria-Pau Ginebra, Salvador Capella-Guitierrez and Osnat Hakimi

Research Track: Short Papers

Keyphrase Extraction from Scientific Articles via Extractive Summarization

Chrysovalantis Giorgos Kontoulis, Eirini Papagiannopoulou and Grigorios Tsoumakas

Argument Mining for Scholarly Document Processing: Taking Stock and Looking Ahead

Khalid Al Khatib, Tirthankar Ghosal, Yufang Hou, Anita de Waard and Dayne Freitag

Bootstrapping Multilingual Metadata Extraction: A Showcase in Cyrillic

Johan Krause, Igor Shapiro, Tarek Saier and Michael Färber

The Effect of Pretraining on Extractive Summarization for Scientific Documents

Yash Gupta, Pawan Sasanka Ammanamanchi, Shikha Bordia, Arjun Manoharan, Deepak Mittal, Ramakanth Pasunuru, Manish Shrivastava, Maneesh Singh, Mohit Bansal and Preethi Jyothi

No Day Set (continued)

Finding Pragmatic Differences Between Disciplines

Lee Kezar and Jay Pujara

Extractive Research Slide Generation Using Windowed Labeling Ranking

Athar Sefid, Prasenjit Mitra, Jian Wu and C Lee Giles

Shared Task: LongSumm

LongSumm 2021: Session based automatic summarization model for scientific document

senci ying, Zheng Yan Zhao and wuhe zou

CNLP-NITS @ LongSumm 2021: TextRank Variant for Generating Long Summaries

Darsh Kaushik, Abdullah Faiz Ur Rahman Khilji, Utkarsh Sinha and Partha Pakray

Unsupervised document summarization using pre-trained sentence embeddings and graph centrality

Juan Ramirez-Orta and Evangelos Milios

Shared Task: SCIVER

QMUL-SDS at SCIVER: Step-by-Step Binary Classification for Scientific Claim Verification

Xia Zeng and Arkaitz Zubiaga