IJCNLP-AACL 2023

The 2nd Workshop on Information Extraction from Scientific Publications

Proceedings of WIESP 2023

November 1, 2023

The IJCNLP-AACL organizers gratefully acknowledge the support from the following sponsors.

Platinum

Gold

Silver

©2023 The Asian Federation of Natural Language Processing and 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 979-8-89176-020-2

2nd WIESP at IJCNLP-AACL 2023

Building on the success of the First WIESP at AACL-IJCNLP 2022, the Second Workshop on Information Extraction from Scientific Publications (WIESP) provided a platform for researchers to foster discussion and research on information extraction, mining, generation, and knowledge discovery from scientific publications using Natural Language Processing and Machine Learning techniques. Much technological change happened in one year (since the 1st WIESP), especially with Generative Artificial Intelligence research. We incorporated a few additional topics to stay abreast with the latest developments and research in the community.

The Workshop on Information Extraction from Scientific Publications (WIESP) is a forum to foster discussion and research using Natural Language Processing and Machine Learning. In this space, leading professionals, organizations, early career researchers and students can cooperate towards building the algorithms, models, and tools that will pave the way for machine comprehension of science in the future.

WIESP 2023 received 22 submissions, of which 17 were accepted (15 papers and 2 shared task system papers).

WIESP 2023 was held on November 1st 2023.

Organizing Committee

Tirthankar Ghosal, Oak Ridge National Laboratory, USA
Felix Grezes, Center for Astrophysics | Harvard & Smithsonian, USA
Thomas Allen, Center for Astrophysics | Harvard & Smithsonian, USA
Sergi Blanco-Cuaresma, Center for Astrophysics | Harvard & Smithsonian, USA; and Laboratoire de Recherche en Neuroimagerie, University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
Alberto Accomazzi, Center for Astrophysics | Harvard & Smithsonian, USA
Kelly Lockhart, Center for Astrophysics | Harvard & Smithsonian, USA

Program Committee

Daniel Acuna
Akiko Aizawa
Thomas Allen
Akhil Pandey Akella
Hamed Alhoori
Atilla Kaan Alkan
Hardik Arora
Premjith B
Arne Binder
Sergi Blanco-Cuaresma
Elena Callegari
Jennifer D'Souza
Vignesh Edithal
Sergey Feldman
Madhusudan Ghosh
Felix Grezes
Hen-Hsen Huang
Daisuke Ikeda
Sarvnaz Karimi
Harksoo Kim
Sandeep Kumar
Asheesh Kumar
Xiangci Li
Kelly Lockhart
Faith Mutinda
Sujit Pal
Ted Pedersen
Rajesh Piryani
Trinita Roy
Atriya Sen
Grishma Sharma
Neil Smalheiser
Wojtek Sylwestrzak
Rohan Tondulkar

Shuntaro Yada
Pierre Zweigenbaum

We thank our program committee members for their critical evaluation and contribution to shape the WIESP 2023 program.

Table of Contents

Investigating the Impact of Syntax-Enriched Transformers on Quantity Extraction in Scientific Texts
Necva Bölücü, Maciej Rybinski and Stephen Wan 1

NanoNER: Named Entity Recognition for Nanobiology Using Experts' Knowledge and Distant Supervision
Ran Cheng, Martin Lentschat and Cyril Labbe 14

Relation Extraction from Scientific Texts in Russian with Limited Training Data
Olga Tikhobaeva and Elena Bruches 25

Extracting Definienda in Mathematical Scholarly Articles with Transformers
Shufan Jiang and Pierre Senellart 31

A Novel Dataset Towards Extracting Virus-Host Interactions
Rasha R. Alshawi, Atriya Sen, Nathan S. Upham and Beckett Sterner 37

Detection of Tortured Phrases in Scientific Literature
Eléna Martel, Martin Lentschat and Cyril Labbe 43

AstroLLaMA: Towards Specialized Foundation Models in Astronomy
Tuan Dung Nguyen, Yuan-Sen Ting, Ioana Ciuca, Charles O'Neill, Ze-Chang Sun, Maja Jabłońska, Sandor Kruk, Ernest Perkowski, Jack Miller, Jason Jason Jingsh Li, Josh Peek, Kartheik Iyer, Tomasz Rozanski, Pranav Khetarpal, Sharaf Zaman, David Brodrick, Sergio J. Rodriguez Mendez, Thang Bui, Alyssa Goodman, Alberto Accomazzi, Jill Naiman, Jesse Cranney, Kevin Schawinski and Roberta Raileanu 49

LaTeX Rainbow: Universal LaTeX to PDF Document Semantic & Layout Annotation Framework
Changxu Duan, Zhiyin Tan and Sabine Bartsch 56

Leveraging the Fusion-in-Decoder for Label Classification
Azumi Okuda, Hideya Mino, Taro Miyazaki and Jun Goto 68

Enhancing Academic Title Generation Using SciBERT and Linguistic Rules
Elena Callegari, Peter Vajdecka, Desara Xhura and Anton Karl Ingason 74

MuLMS: A Multi-Layer Annotated Text Corpus for Information Extraction in the Materials Science Domain
Timo Pierre Schrader, Matteo Finco, Stefan Grünewald, Felix Hildebrand and Annemarie Friedrich 84

An End-to-End Pipeline for Bibliography Extraction from Scientific Articles
Bikash Joshi, Anthi Symeonidou, Syed Mazin Danish and Floris Hermsen 101

Factored Verification: Detecting and Reducing Hallucination in Summaries of Academic Papers
Charlie George and Andreas Stuhlmueller 107

APCS: Towards Argument Based Pros and Cons Summarization of Peer Reviews
Sandeep Kumar, Tirthankar Ghosal and Asif Ekbal 117

On the Use of Language Models for Function Identification of Citations in Scholarly Papers
Tomoki Ikoma and Shigeki Matsubara 130

Automated Citation Function Classification and Context Extraction in Astrophysics: Leveraging Paraphrasing and Question Answering
Hariram Veeramani, Surendrabikram Thapa and Usman Naseem 136

Function of Citation in Astrophysics Literature (FOCAL): Findings of the Shared Task
Felix Grezes, Thomas Allen, Tirthankar Ghosal and Sergi Blanco-Cuaresma 143