AACL-IJCNLP 2022

The 1st Workshop on
Information Extraction from Scientific Publications

Proceedings of the Workshop

November 20, 2022

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 978-1-959429-03-6

Preface

The number of scientific papers published per year has exploded in recent years, strengthening its value as one of the main drivers for scientific progress. In astronomy alone, more than 41,000 new articles are published every year and the vast majority are available either via an open-access model or via preprint services. Indexing the article's full-text in search engines helps discover and retrieve vital scientific information to continue building on the shoulders of giants, informing policy, and making evidence-based decisions. Nevertheless, it is difficult to navigate in this ocean of data; finding articles rely heavily on string matching searches and following citations/references. Still, new approaches are necessary to differentiate the signal from the noise more easily (e.g., finding the key articles about the medical condition we are interested in).

Simple string matching has substantial limitations, human language is ambiguous in nature, context matters, and we frequently use the same word and acronyms to represent a multitude of different meanings. Extracting structured and semantically relevant information from scientific publications (e.g., named-entity recognition, summarization, citation intention, linkage to knowledge graphs) allows better selection and filter articles.

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 received 25 submissions, of which 16 were accepted (8 long papers, 4 short papers, and 4 shared task system papers).

WIESP was held on November 20th 2022.

Organizing Committee

Tirthankar Ghosal, Charles University, CZ
Sergi Blanco-Cuaresma, Center for Astrophysics | Harvard & Smithsonian, USA
Alberto Accomazzi, Center for Astrophysics | Harvard & Smithsonian, USA
Robert M. Patton, Oak Ridge National Laboratory, USA
Felix Grezes, Center for Astrophysics | Harvard & Smithsonian, USA
Thomas Allen, Center for Astrophysics | Harvard & Smithsonian, USA

Program Committee

Min-Yuh Day
Hen-Hsen Huang
Jheng-Long Wu
Daniel Acuna
Akiko Aizawa
Hamed Alhoori
Atilla Kaan Alkan
Thomas Allen
Hardik Arora
Premjith B
Partha Basuchowdhuri
Saprativa Bhattacharjee
Yimeng Dai
Xiang Dai
Vignesh Edithal
Sergey Feldman
Edward Fox
Zheng Gao
Daisuke Ikeda
Sarvnaz Karimi
Valia Kordoni
Sarvnaz Karimi
Valia Kordoni
Asheesh Kumar
Rishu Kumar
Sandeep Kumar
Xiangci Li
Shigeki Matsubara
Yoshitomo Matsubara
Sujit Pal
Rajesh Piryani
Trinita Roy
Neil Smalheiser
Wojtek Sylwestrzak
Rohan Tondulkar
George Tsatsaronis

Jan Philip Wahle
Ronin Wu
Wuhe Zou

Table of Contents

Conference Program

Overview of the First Shared Task on Detecting Entities in the Astrophysics Literature (DEAL)
Felix Grezes, Sergi Blanco-Cuaresma, Thomas Allen and Tirthankar Ghosal

Classification of URL Citations in Scholarly Papers for Promoting Utilization of Research Artifacts
Masaya Tsunokake and Shigeki Matsubara

TELIN: Table Entity LINker for Extracting Leaderboards from Machine Learning Publications
Sean Yang, Chris Tensmeyer and Curtis Wigington

PICO Corpus: A Publicly Available Corpus to Support Automatic Data Extraction from Biomedical Literature
Faith Mutinda, Kongmeng Liew, Shuntaro Yada, Shoko Wakamiya and Eiji ARAMAKI

Linking a Hypothesis Network From the Domain of Invasion Biology to a Corpus of Scientific Abstracts: The INAS Dataset
Marc Brinner, Tina Heger and Sina Zarriess

Leveraging knowledge graphs to update scientific word embeddings using latent semantic imputation
Jason Hoelscher-Obermaier, Edward Stevinson, Valentin Stauber, Ivaylo Zhelev, Viktor Botev, Ronin Wu and Jeremy Minton

Full-Text Argumentation Mining on Scientific Publications
Arne Binder, Leonhard Hennig and Bhuvanesh Verma

On the portability of extractive Question-Answering systems on scientific papers to real-life application scenarios
Chyrine Tahri, Xavier Tannier and Patrick Haouat

Detecting Entities in the Astrophysics Literature: A Comparison of Word-based and Span-based Entity Recognition Methods
Xiang Dai and Sarvnaz Karimi

Domain Specific Augmentations as Low Cost Teachers for Large Students
Po-Wei Huang

Moving beyond word lists: towards abstractive topic labels for human-like topics of scientific documents
Domenic Rosati

Astro-mT5: Entity Extraction from Astrophysics Literature using mT5 Language Model
Madhusudan Ghosh, Payel Santra, Sk Asif Iqbal and Partha Basuchowdhuri

No Day Set (continued)