

# **1st Workshop on Information Extraction from Scientific Publications (WIESP 2022)**

Online  
20 November 2022

ISBN: 978-1-7138-6756-2

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2022) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Overview of the First Shared Task on Detecting Entities in the Astrophysics Literature (DEAL)</i> Felix Grezes, Sergi Blanco-Cuaresma, Thomas Allen and Tirthankar Ghosal .....	1
<i>Classification of URL Citations in Scholarly Papers for Promoting Utilization of Research Artifacts</i> Masaya Tsunokake and Shigeki Matsubara .....	8
<i>TELIN: Table Entity LINKer for Extracting Leaderboards from Machine Learning Publications</i> Sean Yang, Chris Tensmeyer and Curtis Wigington .....	20
<i>PICO Corpus: A Publicly Available Corpus to Support Automatic Data Extraction from Biomedical Literature</i> Faith Mutinda, Kongmeng Liew, Shuntaro Yada, Shoko Wakamiya and Eiji ARAMAKI .....	26
<i>Linking a Hypothesis Network From the Domain of Invasion Biology to a Corpus of Scientific Abstracts: The INAS Dataset</i> Marc Brinner, Tina Heger and Sina Zariess .....	32
<i>Leveraging knowledge graphs to update scientific word embeddings using latent semantic imputation</i> Jason Hoelscher-Obermaier, Edward Stevinson, Valentin Stauber, Ivaylo Zhelev, Viktor Botev, Ronin Wu and Jeremy Minton .....	43
<i>Full-Text Argumentation Mining on Scientific Publications</i> Arne Binder, Leonhard Hennig and Bhuvanesh Verma .....	54
<i>On the portability of extractive Question-Answering systems on scientific papers to real-life application scenarios</i> Chyryne Tahri, Xavier Tannier and Patrick Haouat .....	67
<i>Detecting Entities in the Astrophysics Literature: A Comparison of Word-based and Span-based Entity Recognition Methods</i> Xiang Dai and Sarvnaz Karimi .....	78
<i>Domain Specific Augmentations as Low Cost Teachers for Large Students</i> Po-Wei Huang .....	84
<i>Moving beyond word lists: towards abstractive topic labels for human-like topics of scientific documents</i> Domenic Rosati .....	91
<i>Astro-mT5: Entity Extraction from Astrophysics Literature using mT5 Language Model</i> Madhusudan Ghosh, Payel Santra, Sk Asif Iqbal and Partha Basuchowdhuri .....	100
<i>NLPSharedTasks: A Corpus of Shared Task Overview Papers in Natural Language Processing Domains</i> Anna Martin, Ted Pedersen and Jennifer D’Souza .....	105
<i>Parsing Electronic Theses and Dissertations Using Object Detection</i> Aman Ahuja, Alan Devera and Edward Alan Fox .....	121
<i>TDAC, The First Corpus in Time-Domain Astrophysics: Analysis and First Experiments on Named Entity Recognition</i> Atilla Kaan Alkan, Cyril Grouin, Fabian Schussler and Pierre Zweigenbaum .....	131

*Reproducibility Signals in Science: A preliminary analysis*  
Akhil Pandey Akella, Hamed Alhoori and David Koop ..... 140

*A Majority Voting Strategy of a SciBERT-based Ensemble Models for Detecting Entities in the Astro-  
physics Literature (Shared Task)*  
Atilla Kaan Alkan, Cyril Grouin, Fabian Schussler and Pierre Zweigenbaum ..... 145