

Fourth Workshop on Scholarly Document Processing (SDP 2024)

Bangkok, Thailand
16 August 2024

ISBN: 979-8-3313-0154-5

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© (2024) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

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

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
Web: www.proceedings.com

Table of Contents

<i>Overview of the Fourth Workshop on Scholarly Document Processing</i>	
Tirthankar Ghosal, Amanpreet Singh, Anita De Waard, Philipp Mayr, Aakanksha Naik, Orion Weller, Yoonjoo Lee, Zejiang Shen and Yanxia Qin	1
<i>Overview of the DagPap24 Shared Task on Detecting Automatically Generated Scientific Paper</i>	
Savvas Chamezopoulos, Drahomira Herrmannova, Anita De Waard, Drahomira Herrmannova, Domenic Rosati and Yury Kashnitsky	7
<i>Overview of the Context24 Shared Task on Contextualizing Scientific Claims</i>	
Chu Sern Joel Chan, Aakanksha Naik, Matthew Akamatsu, Hanna Bekele, Erin Bransom, Ian Campbell and Jenna Sparks	12
<i>Controllable Citation Sentence Generation with Language Models</i>	
Nianlong Gu and Richard Hahnloser	22
<i>Toward Related Work Generation with Structure and Novelty Statement</i>	
Kazuya Nishimura, Kuniaki Saito, Tosho Hirasawa and Yoshitaka Ushiku	38
<i>Understanding Survey Paper Taxonomy about Large Language Models via Graph Representation Learning</i>	
Jun Zhuang and Casey Kennington	58
<i>Beyond Retrieval: Topic-based Alignment of Scientific Papers to Research Proposal</i>	
Rudra Nath Palit, Manasi Patwardhan, Lovekesh Vig and Gautam Shroff	70
<i>Evaluating the Effectiveness of Retrieval-Augmented Large Language Models in Scientific Document Reasoning</i>	
Sai Munikoti, Anurag Acharya, Sridevi Wagle and Sameera Horawalavithana	84
<i>Cited Text Spans for Scientific Citation Text Generation</i>	
Xiangci Li, Yi-Hui Lee and Jessica Ouyang	90
<i>CiteAssist: A System for Automated Preprint Citation and BibTeX Generation</i>	
Lars Benedikt Kaesberg, Terry Ruas, Jan Philip Wahle and Bela Gipp	105
<i>An end-to-end entity recognition and disambiguation framework for identifying Author Affiliation from literature publications</i>	
Lianghong Lin, wenxixie-c@my.cityu.edu.hk wenxixie-c@my.cityu.edu.hk, spczili@speed-polyu.edu.hk spczili@speed-polyu.edu.hk and Tianyong Hao	120
<i>Utilizing an Ensemble Model with Anomalous Label Smoothing to Detect Generated Scientific Papers</i>	
Yuan Zhao, Junruo Gao, Junlin Wang, Gang Luo and Liang Tang	130
<i>AffilGood: Building reliable institution name disambiguation tools to improve scientific literature analysis</i>	
Nicolau Duran-Silva, Pablo Accuosto, Piotr Przybyła and Horacio Saggion	135
<i>Metadata Enhancement Using Large Language Models</i>	
Hyunju Song, Steven Bethard and Andrea K. Thomer	145
<i>MISTI: Metadata-Informed Scientific Text and Image Representation through Contrastive Learning</i>	
Pawin Taechoyotin and Daniel Acuna	155

<i>First Steps in Building a Knowledge Base of Mathematical Results</i>	
Shrey Mishra, Yacine Brihmouche, Théo Delemazure, Antoine Gauquier and Pierre Senellart	165
<i>AutoRef : Generating Refinements of Reviews Given Guidelines</i>	
Soham Chitnis, Manasi Patwardhan, ashwin@goa.bits-pilani.ac.in ashwin@goa.bits-pilani.ac.in, tanmayv@goa.bits-pilani.ac.in tanmayv@goa.bits-pilani.ac.in, Lovekesh Vig and Gautam Shroff.	175
<i>Artificial Intuition: Efficient Classification of Scientific Abstracts</i>	
Harsh Sakhrani, Naseela Pervez, Anirudh Ravikumar, Fred Morstatter, Alexandra Graddy-Reed and Andrea Belz	191
<i>Synthetic Context with LLM for Entity Linking from Scientific Tables</i>	
Yuji Oshima, Hiroyuki Shindo, Hiroki Teranishi, Hiroki Ouchi and Taro Watanabe	202
<i>Papilusion at DAGPap24: Paper or Illusion? Detecting AI-generated Scientific Papers</i>	
Nikita Andreev, Alexander Shirnin, Vladislav Mikhailov and Ekaterina Artemova	215
<i>Multi-head Span-based Detector for AI-generated Fragments in Scientific Papers</i>	
German Gritsai, Ildar Khabutdinov and Andrey Grabovoy	220
<i>Guiding Large Language Models via External Attention Prompting for Scientific Extreme Summarization</i>	
Yuan Chang, Ziyue Li and Xiaoqiu Le	226
<i>Simulating Expert Discussions with Multi-agent for Enhanced Scientific Problem Solving</i>	
Ziyue Li, Yuan Chang and Xiaoqiu Le	243
<i>An Analysis of Tasks and Datasets in Peer Reviewing</i>	
Moritz Staudinger, Wojciech Kusa, Florina Piroi and Allan Hanbury	257
<i>Zero-shot Scientific Claim Verification Using LLMs and Citation Text</i>	
Carlos Alvarez, Maxwell Bennett and Lucy Lu Wang	269
<i>Researcher Representations Based on Aggregating Embeddings of Publication Titles: A Case Study in a Japanese Academic Database</i>	
Hiroyoshi Nagao and Marie Katsurai	277
<i>CoSAEmb: Contrastive Section-aware Aspect Embeddings for Scientific Articles</i>	
Shruti Singh and Mayank Singh	283
<i>Integrating Table Representations into Large Language Models for Improved Scholarly Document Comprehension</i>	
Buse Sibel Korkmaz and Antonio Del Rio Chanona	293
<i>Harnessing CLIP for Evidence Identification in Scientific Literature: A Multimodal Approach to Context24 Shared Task</i>	
Anukriti Kumar and Lucy Lu Wang	307
<i>CSIRO at Context24: Contextualising Scientific Figures and Tables in Scientific Literature</i>	
Necva Bölücü, Vincent Nguyen, Roelien C Timmer, Huichen Yang, Maciej Rybinski, Stephen Wan and Sarvnaz Karimi	314
<i>OSX at Context24: How Well Can GPT Tackle Contextualizing Scientific Figures and Tables</i>	
Tosho Hirasawa	324