

COLING

**International Conference on
Computational Linguistics**

Proceedings of the Conference and Workshops

COLING

Volume 29 (2022), No. 5

**Proceedings of Pattern-based Approaches to NLP in the Age
of Deep Learning
(PAN-DL)**

**The 29th International Conference on
Computational Linguistics**

October 12 - 17, 2022
Gyeongju, Republic of Korea

Copyright of each paper stays with the respective authors (or their employers).

ISSN 2951-2093

Message from the Organizers

Welcome to the first edition of the Workshop on Pattern-based Approaches to NLP in the Age of Deep Learning (Pan-DL)! Our workshop is being organized online on October 17, 2022, in conjunction with the 29th International Conference on Computational Linguistics (COLING).

We all know that deep-learning methods have dominated the field of natural language processing in the past decade. However, these approaches usually rely on the availability of high-quality and high-quantity data annotation. Furthermore, the learned models are difficult to interpret and incur substantial technical debt. As a result, these approaches tend to exclude users that lack the necessary machine learning background. In contrast, rule-based methods are easier to deploy and adapt; they support human examination of intermediate representations and reasoning steps; they are more transparent to subject-matter experts; they are amenable to having a human in the loop through intervention, manipulation and incorporation of domain knowledge; and further the resulting systems tend to be lightweight and fast. This workshop focuses on all aspects of rule-based approaches, including their application, representation, and interpretability, as well as their strengths and weaknesses relative to state-of-the-art machine learning approaches.

Considering the large number of potential directions in this neuro-symbolic space, we emphasized inclusivity in our workshop. We received 13 papers and accepted 10 for oral presentation. This resulted in an overall acceptance rate of 77%.

In addition of the oral presentations of the accepted papers, our workshop includes a keynote talk by Ellen Riloff, who has made crucial contributions to the field of natural language processing, many of which are at the intersection of rule- and neural-based methods. Further, the workshop contains a panel that will discuss the merits and limitations of rules in our neural era. The panelists will be academics with expertise in both neural- and rule-based methods, industry experts that employ these methods for commercial products, government officials in charge of AI funding, organizers of natural language processing evaluations, and subject matter experts that have used rule-based methods for domain-specific applications.

We thank Ellen Riloff and the panelists for their important contribution to our workshop!

Finally, we are thankful to the members of the program committee for their insightful reviews! We are confident that all submissions have benefited from their expert feedback. Their contribution was a key factor for accepting a diverse and high-quality list of papers, which we hope will make the first edition of the Pan-DL workshop a success, and will motivate many future editions.

Pan-DL 2022 Organizers
October 2022

Organizing Committee:

Laura Chiticariu, IBM, USA
Yoav Goldberg, Bar Ilan University, Israel
Gus Hahn-Powell, University of Arizona, USA
Clayton T. Morrison, University of Arizona, USA
Rebecca Sharp, Lex Machina, USA
Mihai Surdeanu, University of Arizona, USA
Marco Valenzuela-Escárcega, University of Arizona, USA
Enrique Noriega-Atala, University of Arizona, USA
Aakanksha Naik, Carnegie Mellon University, USA

Program Committee:

Dane Bell, The RiverStone Group, USA
Maria Alexeeva, University of Arizona, USA
Xuan Wang, University of Illinois at Urbana-Champaign, USA
Daniela Claro, Universidade Federal da Bahia, Brazil
Anthony Rios, University of Texas at San Antonio, USA
Aryeh Tiktinsky, Allen Institute for AI (AI2), USA
Juan Diego Rodriguez, University of Texas at Austin, USA
Hillel Taub-Tabib, Allen Institute for AI (AI2), Israel
Roberty Vavareanu, University of Arizona, USA
Andrew Zupon, Google, USA
Hoang N.H Van, University of Arizona, USA

Table of Contents

<i>PatternRank: Jointly Ranking Patterns and Extractions for Relation Extraction Using Graph-Based Algorithms</i>	
Robert Vacareanu, Dane Bell and Mihai Surdeanu	1
<i>Key Information Extraction in Purchase Documents using Deep Learning and Rule-based Corrections</i>	
Roberto Arroyo, Javier Yebes, Elena Martínez, Héctor Corrales and Javier Lorenzo	11
<i>Unsupervised Generation of Long-form Technical Questions from Textbook Metadata using Structured Templates</i>	
Indrajit Bhattacharya, Subhasish Ghosh, Arpita Kundu, Pratik Saini and Tapas Nayak	21
<i>Building Korean Linguistic Resource for NLU Data Generation of Banking App CS Dialog System</i>	
Jeongwoo Yoon, Onyu Park, Changhoe Hwang, Gwanghoon Yoo, Eric Laporte and Jeesun Nam	29
<i>SSP-Based Construction of Evaluation-Annotated Data for Fine-Grained Aspect-Based Sentiment Analysis</i>	
Suwon Choi, Shinwoo Kim, Changhoe Hwang, Gwanghoon Yoo, Eric Laporte and Jeesun Nam	38
<i>Accelerating Human Authorship of Information Extraction Rules</i>	
Dayne Freitag, John Cadigan, John Niekrasz and Robert Sasseen	45
<i>Syntax-driven Data Augmentation for Named Entity Recognition</i>	
Arie Sutiono and Gus Hahn-Powell	56
<i>Query Processing and Optimization for a Custom Retrieval Language</i>	
Yakov Kuzin, Anna Smirnova, Evgeniy Slobodkin and George Chernishev	61
<i>Rule Based Event Extraction for Artificial Social Intelligence</i>	
Remo Nitschke, Yuwei Wang, Chen Chen, Adarsh Pyarelal and Rebecca Sharp	71
<i>Neural-Guided Program Synthesis of Information Extraction Rules Using Self-Supervision</i>	
Enrique Noriega-Atala, Robert Vacareanu, Gus Hahn-Powell and Marco A. Valenzuela-Escárcega	85

