COLING 2025

RegNLP-2025
The First Workshop on Regulatory NLP

Workshop Proceedings

Editors
Tuba Gokhan, Kexin Wang, Iryna Gurevych, Ted Briscoe

January 20, 2025
Abu Dhabi, UAE

©2025 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 979-8-89176-217-6

Preface

Regulatory compliance is a fundamental aspect of governance across industries, yet the complexity and volume of regulatory texts often make adherence a challenging task. The field of Natural Language Processing (NLP) has started to transform this landscape by enabling automated solutions for information retrieval, question answering, and document cross-referencing. Despite this progress, significant challenges remain in adapting general NLP methods to the nuanced and specialized needs of regulatory domains.

The RegNLP-2025 workshop was organized to address these challenges, establish Regulatory NLP as a specialized area, and foster collaboration among researchers, practitioners, and policymakers. This inaugural event provided a platform for exploring advanced methods, tools, and datasets tailored to the regulatory context, aiming to improve the efficiency and accuracy of compliance tasks.

This workshop aimed to bring together NLP researchers, domain experts, and industry stakeholders to discuss developments and challenges in Regulatory NLP. It aimed to introduce new methods for regulatory information retrieval and answer generation while encouraging the exchange of knowledge on the development and evaluation of domain-specific tools and models. The workshop also included the Regulatory Information Retrieval and Answer Generation (RIRAG) shared task, which invited participants to explore innovative approaches for retrieving relevant regulatory passages and generating coherent, accurate answers to compliance-related queries.

We are grateful to the program committee for their detailed reviews, our keynote speakers for their insightful contributions, and the authors and attendees for their participation. Special thanks to the shared task participants for advancing research in regulatory information retrieval and answer generation.

We hope this workshop fosters new collaborations and ideas in the Regulatory NLP domain, paving the way for further advancements in the future.

Tuba Gokhan, Kexin Wang, Iryna Gurevych, Ted Briscoe

January 2025

Organizing Committee

- **Tuba Gokhan**, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI)
- **Kexin Wang**, UKP Lab, Technical University of Darmstadt
- **Iryna Gurevych**, UKP Lab, Technical University of Darmstadt & MBZUAI
- **Ted Briscoe**, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI)

Program Committee

- **Sallam Abualhaija**, University of Luxembourg
- **Chetan Arora**, Monash University
- **Thales Bertaglia**, Maastricht University
- **Travis D. Breaux**, Carnegie Mellon University
- **Silvana Castano**, University of Milan
- **Luigi Di Caro**, University of Turin
- **Ashish Chouhan**, Heidelberg University
- **Chandra Kiran Reddy Evuru**, University of Maryland
- **Alfio Ferrara**, University of Milan
- **Tunga Gungor**, Bogazici University
- **Lena Held**, UKP Lab, Technical University of Darmstadt
- **Timour Igamberdiev**, UKP Lab, Technical University of Darmstadt
- **Daniel Martin Katz**, Chicago-Kent College of Law - Illinois Institute of Technology
- **Manolis Koubarakis**, National and Kapodistrian University of Athens
- **Prodromos Malakasiotis**, Athens University of Economics and Business
- **Luisa Mich**, University of Trento
- **Paulo Quaresma**, University of Evora
- **Carlo Sansone**, University of Naples Federico II
- **Dimitrios Tsarapatsanis**, University of York
- **Peter Vickers**, University of Sheffield
- **Nicola Zeni**, University of Trento

Table of Contents

Shared Task RIRAG-2025: Regulatory Information Retrieval and Answer Generation
Tuba Gokhan, Kexin Wang, Iryna Gurevych and Ted Briscoe 1

Challenges in Technical Regulatory Text Variation Detection
Shriya Vaagdevi Chikati, Samuel Larkin, David Minicola and Chi-kiu Lo 5

Bilingual BSARD: Extending Statutory Article Retrieval to Dutch
Ehsan Lotfi, Nikolay Banar, Nerses Yuzbashyan and Walter Daelemans 10

Unifying Large Language Models and Knowledge Graphs for efficient Regulatory Information Retrieval and Answer Generation
Kishore Vanapalli, Aravind Kilaru, Omair Shafiq and Shahzad Khan 22

A Hybrid Approach to Information Retrieval and Answer Generation for Regulatory Texts
Jhon Stewar Rayo Mosquera, Carlos Raul De La Rosa Peredo and Mario Garrido Cordoba 31

1-800-SHARED-TASKS at RegNLP: Lexical Reranking of Semantic Retrieval (LeSeR) for Regulatory Question Answering
Jebish Purbey, Drishti Sharma, Siddhant Gupta, Khawaja Murad, Siddartha Pullakhandam and Ram Mohan Rao Kadiyala 36

MST-R: Multi-Stage Tuning for Retrieval Systems and Metric Evaluation
Yash Malviya, Karan Dhingra and Maneesh Singh 41

AUEB-Archimedes at RIRAG-2025: Is Obligation concatenation really all you need?
Ioannis Chasandras, Odysseas S. Chlapanis and Ion Androutsopoulos 52

Structured Tender Entities Extraction from Complex Tables with Few-short Learning
Asim Abbas, Mark Lee, Niloofer Shanavas, Venelin Kovatchev and Mubashir Ali 59

A Two-Stage LLM System for Enhanced Regulatory Information Retrieval and Answer Generation
Fengzhao Sun, Jun Yu, Jiaming Hou, yutong lin and Tianyu Liu 68

NUST Nova at RIRAG 2025: A Hybrid Framework for Regulatory Information Retrieval and Question Answering
Mariam Babar Khan, Huma Ameer, Seemab Latif and Mehwish Fatima 73

NUST Alpha at RIRAG 2025: Fusion RAG for Bridging Lexical and Semantic Retrieval and Question Answering
Muhammad Rouhan Faisal, Muhammad Abdullah, Faizyaab Ali Shah, Shalina Riaz, Huma Ameer, Seemab Latif and Mehwish Fatima 79

NUST Omega at RIRAG 2025: Investigating Context-aware Retrieval and Answer Generations-Lessons and Challenges
Huma Ameer, Muhammad Hannan Akram, Seemab Latif and Mehwish Fatima 85

Enhancing Regulatory Compliance Through Automated Retrieval, Reranking, and Answer Generation
Kübranur Umar, Hakan Doğan, Onur Özcan, İsmail Karakaya, Alper Karamanlıoğlu and Berkan Demirel 91

A REGNLP Framework: Developing Retrieval-Augmented Generation for Regulatory Document Analysis

Ozan Bayer, Elif Nehir ULU, Yasemin Sarkın, Ekrem Sütçü, Defne Buse Çelik, Alper Karamanlıoğlu, İsmail Karakaya and Berkan Demirel 97

Regulatory Question-Answering using Generative AI

Devin Quinn, Sumit P. Pai, Iman Yousfi, Nirmala Pudota and Sanmitra Bhattacharya 102

RIRAG: A Bi-Directional Retrieval-Enhanced Framework for Financial Legal QA in ObliQA Shared Task

Xinyan Zhang, Xiaobing Feng, Xiujuan Xu, zhiliang zheng and Kai Wu 107

RAGulator: Effective RAG for Regulatory Question Answering

Islam Aushev, Egor Kratkov, Evgenii Nikoalev, Andrei Vladimirovich Glinskii, Vasilii Krikunov, Alexander Panchenko, Vasily Konovalov and Julia Belikova 114

Conference Program

9:00–10:15 **Session 1**

9:00–9:05 ***Opening Remarks***

9:05–9:35 ***Invited Speaker***

9:35–9:55 *Shared Task RIRAG-2025: Regulatory Information Retrieval and Answer Generation*
Tuba Gokhan, Kexin Wang, Iryna Gurevych and Ted Briscoe

9:55–10:15 *Challenges in Technical Regulatory Text Variation Detection*
Shriya Vaagdevi Chikati, Samuel Larkin, David Minicola and Chi-kiu Lo

10:15–11:00 **Coffee Break**

11:00–12:20 **Session 2**

11:00–11:30 ***Invited Speaker***

11:30–11:55 *Bilingual BSARD: Extending Statutory Article Retrieval to Dutch*
Ehsan Lotfi, Nikolay Banar, Nerses Yuzbashyan and Walter Daelemans

11:55–12:20 *Unifying Large Language Models and Knowledge Graphs for efficient Regulatory Information Retrieval and Answer Generation*
Kishore Vanapalli, Aravind Kilaru, Omair Shafiq and Shahzad Khan

12:20–14:00 Lunch Break

14:00–15:20 Session 3

14:00–14:20 *A Hybrid Approach to Information Retrieval and Answer Generation for Regulatory Texts*
Jhon Stewar Rayo Mosquera, Carlos Raul De La Rosa Peredo and Mario Garrido Cordoba

14:20–14:40 *1-800-SHARED-TASKS at RegNLP: Lexical Reranking of Semantic Retrieval (LeSeR) for Regulatory Question Answering*
Jebish Purbey, Drishti Sharma, Siddhant Gupta, Khawaja Murad, Siddartha Pullakhandam and Ram Mohan Rao Kadiyala

14:40–15:00 *MST-R: Multi-Stage Tuning for Retrieval Systems and Metric Evaluation*
Yash Malviya, Karan Dhingra and Maneesh Singh

15:00–15:20 *AUEB-Archimedes at RIRAG-2025: Is Obligation concatenation really all you need?*
Ioannis Chasandras, Odysseas S. Chlapanis and Ion Androutsopoulos

15:20–16:00 Coffee Break

16:00–17:30 Poster Session

16:00–17:30 *Structured Tender Entities Extraction from Complex Tables with Few-short Learning*
Asim Abbas, Mark Lee, Niloofer Shanavas, Venelin Kovatchev and Mubashir Ali

16:00–17:30 *A Two-Stage LLM System for Enhanced Regulatory Information Retrieval and Answer Generation*
Fengzhao Sun, Jun Yu, Jiaming Hou, yutong lin and Tianyu Liu

16:00–17:30 *NUST Nova at RIRAG 2025: A Hybrid Framework for Regulatory Information Retrieval and Question Answering*
Mariam Babar Khan, Huma Ameer, Seemab Latif and Mehwish Fatima

16:00–17:30 *NUST Alpha at RIRAG 2025: Fusion RAG for Bridging Lexical and Semantic Retrieval and Question Answering*
Muhammad Rouhan Faisal, Muhammad Abdullah, Faizyaab Ali Shah, Shalina Riaz, Huma Ameer, Seemab Latif and Mehwish Fatima

16:00–17:30 *NUST Omega at RIRAG 2025: Investigating Context-aware Retrieval and Answer Generations-Lessons and Challenges*
Huma Ameer, Muhammad Hannan Akram, Seemab Latif and Mehwish Fatima

16:00–17:30 *Enhancing Regulatory Compliance Through Automated Retrieval, Reranking, and Answer Generation*
Kübranur Umar, Hakan Doğan, Onur Özcan, İsmail Karakaya, Alper Karamanlıoğlu and Berkan Demirel

16:00–17:30 *A REGNLP Framework: Developing Retrieval-Augmented Generation for Regulatory Document Analysis*
Ozan Bayer, Elif Nehir ULU, Yasemin Sarkın, Ekrem Sütçü, Defne Buse Çelik, Alper Karamanlıoğlu, İsmail Karakaya and Berkan Demirel

16:00–17:30 *Regulatory Question-Answering using Generative AI*
Devin Quinn, Sumit P. Pai, Iman Yousfi, Nirmala Pudota and Sanmitra Bhattacharya

16:00–17:30 *RIRAG: A Bi-Directional Retrieval-Enhanced Framework for Financial Legal QA in ObliQA Shared Task*
Xinyan Zhang, Xiaobing Feng, Xiujuan Xu, zhiliang zheng and Kai Wu

16:00–17:30 *RAGulator: Effective RAG for Regulatory Question Answering*
Islam Aushev, Egor Kratkov, Evgenii Nikoalev, Andrei Vladimirovich Glinskii, Vasilii Krikunov, Alexander Panchenko, Vasily Konovalov and Julia Belikova