

LIRAI'24: 2nd Workshop on Legal Information Retrieval meets Artificial Intelligence

Sabine Wehnert
sabine.wehnert@gei.de
Otto von Guericke University
Magdeburg, Germany
Leibniz Institute for Educational
Media | Georg Eckert Institute
Brunswick, Germany

Manuel Fiorelli
manuel.fiorelli@uniroma2.it
Tor Vergata University of Rome
Rome, Italy

Davide Picca
davide.picca@unil.ch
University of Lausanne
Lausanne, Switzerland

Ernesto William De Luca
deluca@gei.de
Otto von Guericke University
Magdeburg, Germany
Leibniz Institute for Educational
Media | Georg Eckert Institute
Brunswick, Germany

Armando Stellato
stellato@uniroma2.it
Tor Vergata University of Rome
Rome, Italy

ABSTRACT

LIRAI is a workshop series on Legal Information Retrieval and Legal Artificial Intelligence. It provides a forum for discussing current trends and challenges in legal artificial intelligence, specifically related to the hypertext nature of legal documents and retrieval tasks. The second edition of LIRAI focuses on three main directions: explainable / justifiable artificial intelligence, hybrid systems that combine formal approaches and machine learning-based methods, including deep learning-based methods, and finally generative artificial intelligence. We call for contributions on these topics in the form of short and long papers, and we aim to publish them as open-access proceedings on CEUR-WS.org once again.

CCS CONCEPTS

• **Applied computing** → Law; • **Information systems** → Information retrieval; • **Computing methodologies** → Artificial intelligence.

KEYWORDS

Legal Informatics, Legal Information Retrieval, Legal Knowledge Representation, Legal Text Mining, Legal Compliance, FAIRness, Semantic Web, Linguistic Legal Linked Open Data, Explainable AI, High-Recall Retrieval, Hybrid Approaches, Generative AI

ACM Reference Format:

Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato. 2024. LIRAI'24: 2nd Workshop on Legal Information Retrieval meets Artificial Intelligence. In *35th ACM Conference on Hypertext*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
HT '24, September 10–13, 2024, Poznan, Poland
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0595-3/24/09.
<https://doi.org/10.1145/3648188.3675120>

and Social Media (HT '24), September 10–13, 2024, Poznan, Poland. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3648188.3675120>

1 DESCRIPTION OF THE WORKSHOP

The Legal Information Retrieval meets Artificial Intelligence (LIRAI) workshop serves as a forum for exploring the interrelation of legal information retrieval and artificial intelligence. Building on the success of the inaugural edition, this year's workshop continues to foster collaboration among researchers, practitioners, and enthusiasts to unravel the intricate relationship between legal research and cutting-edge AI advancements. We anticipate a workshop filled with stimulating discussions on innovative ideas to overcome challenges posed by the complicated nature of legal text.

In the realm of legal documents, hypertext facilitates dynamic linkage between statutes, cases, regulations, and other legal resources. By embedding hyperlinks within the text, users can seamlessly navigate between related sections, cross-referenced materials, or even external sources. This interconnected approach mirrors the inherent relationships within the legal landscape, allowing practitioners and researchers to follow logical connections and trace the evolution of legal principles.

The core topics of this workshop are similar to those of the previous edition, and appended by the current foci:

- Hypertext-based legal systems (e.g., [3, 6, 12])
- Legal information extraction / retrieval (e.g., [2, 8, 11, 18, 24])
- Legal knowledge graphs / ontologies (e.g., [14])
- Relation extraction from legal documents (e.g., [4, 21])
- Explainability / Justifiability in legal retrieval (e.g., [20])
- High-recall settings in legal document retrieval (e.g., [5, 13])
- Legal document formats and organization (e.g., [1, 16, 19])
- FAIR [25] publication of legal documents (e.g., [9])
- Hybrid systems of formal approaches and machine learning or deep learning for legal retrieval (e.g., [17])
- Generative artificial intelligence in legal information extraction / retrieval (e.g., [10, 15])

2 RELEVANCE OF THE WORKSHOP TO THE HYPERTEXT COMMUNITIES

There is a long-standing relationship between legal documents and hypertext for legal information management. The capabilities of hypertext technologies unlock new dimensions of accessibility, collaboration, and efficiency in navigating the intricacies of legal information. Working with legal documents means navigating information about many different kinds of relationships, including explicit connections through citations, implicit ties arising from shared concepts, and hierarchical structures like national laws influenced by international agreements. Amendments contribute to dynamic relationships, while interconnected regulations and guidelines provide supplementary guidance. Precedential relationships emerge from case law, and cross-referencing within documents ensures coherence. Parallel legal systems, present in federal setups, denote relationships between documents at different jurisdictional levels (e.g., federal and state laws). Recognizing these relationships is crucial for comprehensive legal analysis, guiding practitioners through the intricate web of legal knowledge.

3 WORKSHOP ORGANISERS' BIOS

Sabine Wehnert, M.Sc., a Ph.D. candidate at Otto von Guericke University Magdeburg, focuses on legal retrieval, information extraction, explainable AI, knowledge graphs, and usability. Aside from coordinating a Usability Lab at the Leibniz-Institute, her dissertation develops HONto, a knowledge graph from scientific textbooks for legal retrieval and recommendation tasks. She won the Statute Norm Retrieval task in the 2021 COLIEE competition and has become a program committee member since then.

Manuel Fiorelli, Ph.D. is a Research Fellow at Tor Vergata University of Rome, specializing in knowledge engineering and semantic technologies for the Semantic Web. Author of over 30 publications, he has been a PC member of the MTSR conference since 2022, and he is a member of the W3C Ontology-Lexica Community Group. In Legal Informatics, he contributed to LegalHTML, being adopted as a dissemination format for EUR-Lex portal of the EU legislation. Dr. Fiorelli has contributed to R&D projects funded by the DIGITAL program, and he participated in the EU-funded projects SEMAGROW and KATY.

Davide Picca, Ph.D. is an established Digital Humanities researcher with a focus on cultural heritage and digital technologies. His research delves into computational semantics and ontology, exploring the impact of the legal domain on societies and cultures. Driven by a dedication to preserving cultural heritage in the digital era, his contributions aim to advance meaningful interdisciplinary research.

Armando Stellato, Ph.D., an Associate Professor at Tor Vergata University of Rome, is an expert in Knowledge Engineering and Knowledge-Based Systems. With over 100 publications, he actively contributes to international conferences and workshops in the Semantic Web and Natural Language Processing. Engaged in numerous EU-funded projects, including the W3C Ontology-Lexica Community Group, Dr. Stellato collaborates with institutions like FAO, ESA, UN, USDA, and governments, providing expertise in data and documental archives. Leading DIGITAL program-funded projects, he oversees the development of a knowledge-management

ecosystem with platforms like VocBench. In Legal Informatics, he collaborates with the Italian government for semantic organization of laws, and contributed to LegalHTML, a semantic representation model for legal acts adopted by the EU Publications Office.

Ernesto William De Luca leads the Human-Centred Technologies for Educational Media department at the Georg Eckert Institute. Since October 2019, he is a Full Professor in Human-Centred Artificial Intelligence at Otto von Guericke University Magdeburg. Appointed as an associate professor in "computational engineering" in May 2015 by Guglielmo Marconi University, Prof. De Luca's academic journey began with a focus on computational linguistics, leading to a doctorate in computer science. His prolific research spans AI, Machine Learning, Natural Language Processing, Digital Humanities, Semantic Web, and Information Retrieval, with over 200 papers. Actively contributing to conferences and journals, he organizes events and serves as a reviewer for esteemed publications.

4 MOTIVATION

LIRAI is relevant to the hypertext community because attendees can explore the practical implications of AI in legal technology, gain insight into domain-specific considerations of explainability, high-recall settings, and understand potential applications within hypertext systems. The workshop encourages interdisciplinary collaboration at the intersection of hypertext technology and law.

5 WORKSHOP AND SUBMISSION FORMATS

LIRAI features paper presentations and a keynote by a distinguished researcher. Each presentation has a 20-minute slot, allowing 10-15 minutes for the talk and 5-10 minutes for discussion, with potential extensions for particularly engaging topics. Talks may anchor discussions for audience interaction. All submitted papers undergo a single-blind review. The previous reviewers¹ and authors of accepted papers [2, 3, 8, 11, 18, 24] for LIRAI 2023 will be invited into the new program committee. LIRAI aims to publish proceedings through CEUR-WS.org, but the final decision rests with the publisher. Promotion involves announcements on mailing lists and social networks, a dedicated website, email calls for papers, and advertising at related events. The anticipated audience size is approximately 30-40 participants. LIRAI is scheduled for half a day but can become a full-day event depending on the conference's needs and accepted submissions.

6 PREVIOUS EDITIONS OF THE WORKSHOP SERIES

The first edition of LIRAI²[7] was held at the ACM Hypertext Conference 2023³ in Rome, Italy. There were 6 accepted papers out of 8 submissions, complemented by a compelling keynote address delivered by Enrico Francesconi on "Profiles of Knowledge Representation and Reasoning for Legal Information Retrieval and Compliance Checking". The engaging sessions sparked vibrant discussions among an audience of about 20 participants, which were managed within a hybrid setting and were summarized by the organizers [23] in the CEUR-WS.org proceedings [22].

¹<https://sites.google.com/view/lirai-2023/program-committee>

²<https://sites.google.com/view/lirai-2023/>

³<https://ht.acm.org/ht2023/>

REFERENCES

- [1] [n. d.]. *CEN MetaLex: Open XML Interchange Format for Legal and Legislative Resources*. Retrieved July 24, 2023 from <http://www.metalex.eu/>
- [2] Emmanuel Bauer, Dominik Stambach, Nianlong Gu, and Elliott Ash. 2023. Legal Extractive Summarization of U.S. Court Opinions. In *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023 (CEUR Workshop Proceedings, Vol. 3594)*, Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). CEUR-WS.org, 10–30. <https://ceur-ws.org/Vol-3594/paper1.pdf>
- [3] Daniele Bertillo, Andrea de Donato, Carlo Marchetti, and Paolo Meriardo. 2023. Enhancing Accessibility of Parliamentary Video Streams: AI-Based Automatic Indexing Using Verbatim Reports. In *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023 (CEUR Workshop Proceedings, Vol. 3594)*, Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). CEUR-WS.org, 31–40. <https://ceur-ws.org/Vol-3594/paper2.pdf>
- [4] Yanguang Chen, Yuanyuan Sun, Zhihao Yang, and Hongfei Lin. 2020. Joint Entity and Relation Extraction for Legal Documents with Legal Feature Enhancement. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8–13, 2020*, Donia Scott, N ria Bel, and Chengqing Zong (Eds.). International Committee on Computational Linguistics, 1561–1571. <https://doi.org/10.18653/V1/2020.COLING-MAIN.137>
- [5] Charles Courchaine, Tasnova Tabassum, Corey Wade, and Ricky J. Sethi. 2023. Explainable e-Discovery (XeD) Using an Interpretable Fuzzy ARTMAP Neural Network for Technology-Assisted Review. In *IEEE International Conference on Big Data, BigData 2023, Sorrento, Italy, December 15–18, 2023*, Jingrui He, Themis Palpanas, Xiaohua Hu, Alfredo Cuzzocrea, Dejeng Dou, Dominik Slezak, Wei Wang, Aleksandra Gruca, Jerry Chun-Wei Lin, and Rakesh Agrawal (Eds.). IEEE, 2761–2766. <https://doi.org/10.1109/BigData59044.2023.10386391>
- [6] Wojciech Cyrul and Tomasz Pelech-Pilichowski. 2020. Legislating in hypertext. *Opolskie Studia Administracyjne-Prawne* 18, 2 (2020).
- [7] Ernesto William De Luca, Manuel Fiorelli, Davide Picca, Armando Stellato, and Sabine Wehnert. 2023. Legal Information Retrieval meets Artificial Intelligence (LIRAI). In *Proceedings of the 34th ACM Conference on Hypertext and Social Media (Rome, Italy) (HT '23)*. Association for Computing Machinery, New York, NY, USA, Article 46, 4 pages. <https://doi.org/10.1145/3603163.3610575>
- [8] Rima Dessi, Hidir Aras, and Lei Zhang. 2023. DeepKEA: Employing Deep Learning Models for Keyword Extraction from Patent Documents. In *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023 (CEUR Workshop Proceedings, Vol. 3594)*, Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). CEUR-WS.org, 41–46. <https://ceur-ws.org/Vol-3594/paper3.pdf>
- [9] Enrico Francesconi. 2006. The “Norme in Rete” project: Standards and Tools for Italian Legislation. *International Journal of Legal Information* 34, 2 (2006), 358–376. <https://doi.org/10.1017/S0731126500001517>
- [10] Randy Goebel, Yoshinobu Kano, Mi-Young Kim, Julian Rabelo, Ken Satoh, and Masaharu Yoshioka. 2024. Overview and Discussion of the Competition on Legal Information, Extraction/Entailment (COLIEE) 2023. *The Review of Socionetwork Strategies* (2024), 1–21.
- [11] Candida Maria Greco and Andrea Tagarelli. 2023. Topic Similarities in Rights and Duties across European Constitutions using Transformer-based Language Models. In *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023 (CEUR Workshop Proceedings, Vol. 3594)*, Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). CEUR-WS.org, 47–62. <https://ceur-ws.org/Vol-3594/paper4.pdf>
- [12] Graham Greenleaf, Andrew Mowbray, and Philip Chung. 2018. Building sustainable free legal advisory systems: Experiences from the history of AI & law. *Comput. Law Secur. Rev.* 34, 2 (2018), 314–326. <https://doi.org/10.1016/J.CLSR.2018.02.007>
- [13] Timo Kats, Peter van der Putten, and Jan Scholtes. 2023. Relevance feedback strategies for recall-oriented neural information retrieval. *CoRR abs/2311.15110* (2023). <https://doi.org/10.48550/ARXIV.2311.15110> arXiv:2311.15110
- [14] Juli n Moreno Schneider, Georg Rehm, Elena Montiel-Ponsoda, Victor Rodr guez-Doncel, Patricia Mart n-Chozas, Mar a Navas-Loro, Martin Kaltenb ck, Artem Revenko, Sotirios Karamatakis, Christian Sageder, Jorge Gracia, Filippo Maganza, Ilan Kernerman, Dorielle Lonke, Andis Lagzdins, Julia Bosque Gil, Pieter Verhoeven, Elsa Gomez Diaz, and Pascual Boil Ballesteros. 2022. Lynx: A Knowledge-Based AI Service Platform for Content Processing, Enrichment and Analysis for the Legal Domain. *Inf. Syst.* 106, C (may 2022), 18 pages. <https://doi.org/10.1016/j.is.2021.101966>
- [15] Chau Nguyen, Phuong Nguyen, Thanh Tran, Dat Nguyen, An Trieu, Tin Pham, Anh Dang, and Le-Minh Nguyen. 2024. Captain at coliee 2023: Efficient methods for legal information retrieval and entailment tasks. *arXiv preprint arXiv:2401.03551* (2024).
- [16] Monica Palmirani and Fabio Vitali. 2011. Akoma-Ntoso for Legal Documents. In *Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management*, Giovanni Sartor, Monica Palmirani, Enrico Francesconi, and Maria Angela Biasiotti (Eds.). Springer Netherlands, Dordrecht, 75–100. https://doi.org/10.1007/978-94-007-1887-6_6
- [17] Guilherme Paulino-Passos and Francesca Toni. 2023. Learning Case Relevance in Case-Based Reasoning with Abstract Argumentation. In *Legal Knowledge and Information Systems - JURIX 2023: The Thirty-sixth Annual Conference, Maastricht, The Netherlands, 18-20 December 2023 (Frontiers in Artificial Intelligence and Applications, Vol. 379)*, Giovanni Sileno, Jerry Spanakis, and Gijs van Dijk (Eds.). IOS Press, 95–100. <https://doi.org/10.3233/FAIA230950>
- [18] Andrea Simeri and Andrea Tagarelli. 2023. GDPR Article Retrieval based on Domain-adaptive and Task-adaptive Legal Pre-trained Language Models. In *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023 (CEUR Workshop Proceedings, Vol. 3594)*, Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). CEUR-WS.org, 63–76. <https://ceur-ws.org/Vol-3594/paper5.pdf>
- [19] Armando Stellato and Manuel Fiorelli. 2023. LegalHTML: A Representation Language for Legal Acts. In *The Semantic Web: 20th International Conference, ESWC 2023, Heraklion, Crete, Greece, May 28–June 1, 2023, Proceedings* (Heraklion, Greece). Springer-Verlag, Berlin, Heidelberg, 520–537. https://doi.org/10.1007/978-3-031-33455-9_31
- [20] Sabine Wehnert. 2023. Justifiable Artificial Intelligence: Engineering Large Language Models for Legal Applications. *CoRR abs/2311.15716* (2023). <https://doi.org/10.48550/ARXIV.2311.15716> arXiv:2311.15716
- [21] Sabine Wehnert and Ernesto William De Luca. 2021. Finding Implicit Links Between Norms Using HONto. In *Proceedings of the First International Workshop RELATED - Relations in the Legal Domain 2021 co-located with the 18th International Conference on Artificial Intelligence and Law (ICAIL 2021) (CEUR Workshop Proceedings, Vol. 2896)*, Emilio Sulis, Llio Humphreys, Valentina Leone, and Maria Angela Amantea (Eds.). CEUR-WS.org, 47–62. https://ceur-ws.org/Vol-2896/RELATED_2021_paper_4.pdf
- [22] Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). 2023. *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023*. CEUR Workshop Proceedings, Vol. 3594. CEUR-WS.org. <https://ceur-ws.org/Vol-3594>
- [23] Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato. 2023. Summary of the Workshop Legal Information Retrieval meets Artificial Intelligence (LIRAI'23). In *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023 (CEUR Workshop Proceedings, Vol. 3594)*, Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). CEUR-WS.org, 1–8. <https://ceur-ws.org/Vol-3594/paper0.pdf>
- [24] Sabine Wehnert, Davide Picca, and Ernesto William De Luca. 2023. Mining Sentiment and Subjectivity in Swiss Case Law. In *Proceedings of the 1st Legal Information Retrieval meets Artificial Intelligence Workshop LIRAI 2023 co-located with the 34th ACM Hypertext Conference HT 2023, Rome, Italy, September 04, 2023 (CEUR Workshop Proceedings, Vol. 3594)*, Sabine Wehnert, Manuel Fiorelli, Davide Picca, Ernesto William De Luca, and Armando Stellato (Eds.). CEUR-WS.org, 77–90. <https://ceur-ws.org/Vol-3594/paper6.pdf>
- [25] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Merc  Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (15 Mar 2016), 160018. <https://doi.org/10.1038/sdata.2016.18>