

NLLP 2022

Natural Legal Language Processing Workshop 2022

Proceedings of the Workshop

December 8, 2022

The NLLP organizers gratefully acknowledge the support from the following sponsors.

Gold



LBox



©2022 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 978-1-959429-18-0

Introduction

Welcome to the fourth edition of the NLLP (Natural Legal Language Processing) Workshop, co-located with the 2022 Conference on Empirical Methods in Natural Language Processing.

Different industrial sectors have embraced natural language processing (NLP) technologies, which have altered services and products in healthcare, finance, education among others. The legal domain provides enormous potential for generating interesting research problems. Electronic tools are increasingly used for all types of legal tasks and that use is predicted to grow sharply. By its very nature, the practice of law necessarily involves the analysis and interpretation of language. The potential for NLP applications to provide benefit to practitioners of law and consumers of legal services around the world is enormous. We organized this workshop to bring together researchers and practitioners from around the world who develop NLP techniques for legal data. This is an exciting opportunity to expand the boundaries of our field by identifying new problems and exploring new data as it interacts with the full inventory of NLP and machine learning approaches. In this spirit, the Organizing and Program Committee was assembled to include researchers from both academia and industry, from NLP and legal backgrounds.

We were interested in the following types of papers: (1) applications of NLP methods to legal tasks; (2) experimental results using and adapting NLP methods in legal documents; (3) descriptions of new legal tasks for NLP; (4) creation of curated and/or annotated resources; (5) descriptions of systems which use NLP technologies for legal text; (6) industrial research in this area and (7) interdisciplinary position papers.

We again received a record number of submissions. Out of 56 submissions, we accepted 33 papers for an overall acceptance rate of 58.9% percent. Out of the 33 accepted papers, 25 were long and 8 are short. These also include 3 original papers submitted as non-archival, in order to accommodate publication of the work at a later date in a conference or journal. All papers were reviewed by at least 3 members of the Program Committee. In addition, we also offered a venue for presentation for 6 papers accepted to the Findings of EMNLP 2022 on the above topics. All papers were invited to have an oral presentation and in-person attendees are also invited to join a poster session over lunch.

The papers cover a wide range of topics including new data sets for legal NLP, demos, legal perspectives on NLP topics and data rights, methods for dealing with legal documents including processing of long documents, domain adaptation, privacy-aware models and active learning as well as applications of NLP tasks to legal documents including retrieval, information extraction, simplification, extractive and abstractive summarization, generation, named entity recognition, segmentation, document similarity, stance detection and argument reasoning.

We thank our invited speaker Professor Michael A. Livermore from the University of Virginia School of Law for accepting our invitation. In the tradition of past NLLP workshops, the invited speaker is a legal scholar with an interest in empirical methods for legal analysis including NLP methods. For this edition, we are also hosting a panel From NLLP to legal NLP: The Future of the Field.

We thank everyone who expressed interest in the workshop, all authors of submitted papers, members of the Program Committee who did an excellent job at reviewing papers given a short turnaround time, everyone attending the workshop, the EMNLP 2022 conference for hosting us and the workshop and publication chairs for their support. We thank our sponsors – LBox, Bloomberg and the European Research Council Starting Grant project HUMANads – for their contributions.

The NLLP Workshop organizers.

<http://nllpw.org/workshop>

Organizing Committee

Organizers

Nikolaos Aletras, The University of Sheffield

Ilias Chalkidis, University of Copenhagen

Leslie Barrett, Bloomberg Law

Cătălina Goanță, Utrecht University

Daniel Preoțiuc-Pietro, Bloomberg

Program Committee

Reviewers

Tomaso Agnoloni, Ion Androutsopoulos, Elliott Ash

Tim Baldwin, Andrew Blair-Stanek, Lukasz Borchmann, Baldwin Breck

Rajarithnam Chandramouli, Daniel Chen, Ashish Chouhan, Corinna Coupette

Marina Danilevsky, Tony Davis, Stefania Degaetano-Ortlieb, Luigi Di Caro, Kasper Drazewski, Arthur Dyevre

Emmanouil Fergadiotis

Frank Giaoui, Matthias Grabmair, Neel Guha

Ivan Habernal, Peter Henderson, Nils Holzenberger

Daniel Katz, Ilan Kernerman, Manolis Koubarakis

Junyi Jessy Li, Ruta Liepina, How Khang Lim, Antoine Louis

Prodromos Malakasiotis, Adam Meyers, Jelena Mitrovic

Pietro Ortolani

Paulo Quaresma

Georg Rehm

George Sanchez, Cristiana Santos, Thibault Schrepel, Sebastian Schwemer, Mathias Siems, Dan Simonson, Jerrold Soh, Gerasimos Spanakis

Andrea Tagarelli, Dimitrios Tsarapatsanis

Josef Valvoda, Gijs Van Dijck, Jacob van de Kerkhof

Adam Wyner

Marcos Zampieri, Miri Zilka

Keynote Talk: Finding the Law

Mike Livermore

University of Virginia School of Law

Abstract: This presentation will examine challenges in finding the law, for both legal practitioners and scholars engaged in the computational analysis of law. For the practitioner, the challenge is one of search, a process that can be modeled and studied. Although undertheorized, law search has substantial jurisprudential and practical consequences that are only beginning to be explored. For the computational scholar, challenges of selection and data bias are pervasive, and credible scholarship must ground descriptive and causal analyses in the actual processes that generate the data available for study.

Bio: Michael A. Livermore is a Professor of Law at the University of Virginia. He is one of the early scholars involved in a new research paradigm in legal scholarship that uses computational text analysis tools to study the law and legal institutions. Livermore is the author of dozens of academic works, which have appeared in top law journals as well as peer-reviewed legal, scientific, and social science journals. With Daniel N. Rockmore, he edited *Law as Data: Computation, Text, and the Future of Legal Analysis* (Santa Fe Institute Press, 2019). Livermore hosts the Online Workshop on the Computational Analysis of Law, a global forum for scholars to present cutting-edge research in this area. Livermore is also a leading expert on the use of cost-benefit analysis to evaluate regulation. Prior to joining the faculty, Livermore was the founding executive director of the Institute for Policy Integrity at New York University School of Law, a think tank dedicated to improving the quality of government decision-making. He is a public member of the Administrative Conference of the United States.

Keynote Talk: From NLLP to legal NLP - The Future of the Field

Panel

Table of Contents

<i>On Breadth Alone: Improving the Precision of Terminology Extraction Systems on Patent Corpora</i> Sean Nordquist and Adam Meyers	1
<i>On What it Means to Pay Your Fair Share: Towards Automatically Mapping Different Conceptions of Tax Justice in Legal Research Literature</i> Reto Gubelmann, Peter Hongler, Elina Margadant and Siegfried Handschuh	12
<i>ClassActionPrediction: A Challenging Benchmark for Legal Judgment Prediction of Class Action Cases in the US</i> Gil Semo, Dor Bernsohn, Ben Hagag, Gila Hayat and Joel Niklaus	31
<i>Combining WordNet and Word Embeddings in Data Augmentation for Legal Texts</i> Sezen Perçin, Andrea Galassi, Francesca Lagioia, Federico Ruggeri, Piera Santin, Giovanni Sartor and Paolo Torroni	47
<i>A Legal Approach to Hate Speech – Operationalizing the EU’s Legal Framework against the Expression of Hatred as an NLP Task</i> Frederike Zufall, Marius Hamacher, Katharina Kloppenborg and Torsten Zesch	53
<i>Privacy Pitfalls of Online Service Terms and Conditions: a Hybrid Approach for Classification and Summarization</i> Emilia Lukose, Suparna De and Jon Johnson	65
<i>Abstractive Summarization of Dutch Court Verdicts Using Sequence-to-sequence Models</i> Marijn Schraagen, Floris Bex, Nick Van De Luijtgarden and Daniël Prijs	76
<i>Legal-Tech Open Diaries: Lesson learned on how to develop and deploy light-weight models in the era of humongous Language Models</i> Stelios Maroudas, Sotiris Legkas, Prodromos Malakasiotis and Ilias Chalkidis	88
<i>Towards Cross-Domain Transferability of Text Generation Models for Legal Text</i> Vinayshkhar Bannihatti Kumar, Kasturi Bhattacharjee and Rashmi Gangadharaiyah	111
<i>Parameter-Efficient Legal Domain Adaptation</i> Jonathan Li, Rohan Bhambhoria and Xiaodan Zhu	119
<i>Processing Long Legal Documents with Pre-trained Transformers: Modding LegalBERT and Longformer</i> Dimitris Mamakas, Petros Tsotsi, Ion Androutsopoulos and Ilias Chalkidis	130
<i>Data-efficient end-to-end Information Extraction for Statistical Legal Analysis</i> Wonseok Hwang, Saehee Eom, Hanuhl Lee, Hai Jin Park and Minjoon Seo	143
<i>Semantic Segmentation of Legal Documents via Rhetorical Roles</i> Vijit Malik, Rishabh Sanjay, Shouvik Kumar Guha, Angshuman Hazarika, Shubham Nigam, Arnab Bhattacharya and Ashutosh Modi	153
<i>Privacy-Preserving Models for Legal Natural Language Processing</i> Ying Yin and Ivan Habernal	172
<i>Named Entity Recognition in Indian court judgments</i> Prathamesh Kalamkar, Astha Agarwal, Aman Tiwari, Smita Gupta, Saurabh Karn and Vivek Raghavan	184

<i>The Legal Argument Reasoning Task in Civil Procedure</i> Leonard Bongard, Lena Held and Ivan Habernal	194
<i>Efficient Deep Learning-based Sentence Boundary Detection in Legal Text</i> Reshma Sheik, Gokul T and S Nirmala	208
<i>Tracking Semantic Shifts in German Court Decisions with Diachronic Word Embeddings</i> Daniel Braun	218
<i>Should I disclose my dataset? Caveats between reproducibility and individual data rights</i> Raysa M. Benatti, Camila M. L. Villarroel, Sandra Avila, Esther L. Colombini and Fabiana Severi	228
<i>Attack on Unfair ToS Clause Detection: A Case Study using Universal Adversarial Triggers</i> Shanshan Xu, Irina Broda, Rashid Haddad, Marco Negrini and Matthias Grabmair	238
<i>E-NER — An Annotated Named Entity Recognition Corpus of Legal Text</i> Ting Wai Terence Au, Vasileios Lampos and Ingemar Cox	246
<i>Detecting Relevant Differences Between Similar Legal Texts</i> Xiang Li, Jiaxun Gao, Diana Inkpen and Wolfgang Alschner	256
<i>Legal and Political Stance Detection of SCOTUS Language</i> Noah Bergam, Emily Allaway and Kathleen Mckeown	265
<i>Graph-based Keyword Planning for Legal Clause Generation from Topics</i> Sagar Joshi, Sumanth Balaji, Aparna Garimella and Vasudeva Varma	276
<i>Automatic Classification of Legal Violations in Cookie Banner Texts</i> Marieke Van Hofslot, Almila Akdag Salah, Albert Gatt and Cristiana Santos	287
<i>Text Simplification for Legal Domain: {I}nsights and Challenges</i> Aparna Garimella, Abhilasha Sancheti, Vinay Aggarwal, Ananya Ganesh, Niyati Chhaya and Nandakishore Kambhatla	296
<i>Legal Named Entity Recognition with Multi-Task Domain Adaptation</i> Răzvan-alexandru Smădu, Ion-robert Dinică, Andrei-marius Avram, Dumitru-clementin Cercel, Florin Pop and Mihaela-claudia Cercel	305
<i>Computing and Exploiting Document Structure to Improve Unsupervised Extractive Summarization of Legal Case Decisions</i> Yang Zhong and Diane Litman	322
<i>AraLegal-BERT: A pretrained language model for Arabic Legal text</i> Muhammad Al-qurishi, Sarah Alqaseemi and Riad Souissi	338
<i>An Efficient Active Learning Pipeline for Legal Text Classification</i> Sepideh Mamooler, Rémi Lebret, Stephane Massonnet and Karl Aberer	345

Program

Thursday, December 8, 2022

09:00 - 10:40 *Session 1*

Multi-LexSum: Real-world Summaries of Civil Rights Lawsuits at Multiple Granularities

Zejiang Shen, Kyle Lo, Lauren Yu, Nathan Dahlberg, Margo Schlanger and Doug Downey

Extractive Summarization of Legal Decisions using Multi-task Learning and Maximal Marginal Relevance

Matthias Grabmair, Shanshan Xu and Abhishek Agarwal

Towards Cross-Domain Transferability of Text Generation Models for Legal Text

Vinayshekhar Bannihatti Kumar, Kasturi Bhattacharjee and Rashmi Gangadharaiyah

Parameter-Efficient Legal Domain Adaptation

Jonathan Li, Rohan Bhambhoria and Xiaodan Zhu

ClassActionPrediction: A Challenging Benchmark for Legal Judgment Prediction of Class Action Cases in the US

Gil Semo, Dor Bernsohn, Ben Hagag, Gila Hayat and Joel Niklaus

Zero Shot Transfer of Legal Judgement Prediction as Article-aware Entailment for the European Court of Human Rights

T.y.s.s Santosh, Oana Ichim and Matthias Grabmair

Revisiting Transformer-based Models for Long Document Classification

Desmond Elliott, Sune Darkner, Ilias Chalkidis and Xiang Dai

Attack on Unfair ToS Clause Detection: A Case Study using Universal Adversarial Triggers

Shanshan Xu, Irina Broda, Rashid Haddad, Marco Negrini and Matthias Grabmair

AraLegal-BERT: A pretrained language model for Arabic Legal text

Muhammad Al-qurishi, Sarah Alqaseemi and Riad Souissi

An Efficient Active Learning Pipeline for Legal Text Classification

Sepideh Mamooler, Rémi Lebret, Stephane Massonnet and Karl Aberer

Thursday, December 8, 2022 (continued)

A Legal Approach to Hate Speech – Operationalizing the EU’s Legal Framework against the Expression of Hatred as an NLP Task

Frederike Zufall, Marius Hamacher, Katharina Kloppenborg and Torsten Zesch

Validity Assessment of Legal Will Statements as Natural Language Inference

Mihai Surdeanu, Derek Bambauer, Clayton Morrison, Jacob Israelsen and Alice Kwak

10:40 - 11:00 *Break*

11:00 - 12:00 *Finding the Law - Michael A. Livermore (University of Virginia School of Law)*

12:00 - 12:40 *Session 2*

Data-efficient end-to-end Information Extraction for Statistical Legal Analysis

Wonseok Hwang, Saehee Eom, Hanuhl Lee, Hai Jin Park and Minjoon Seo

Efficient Deep Learning-based Sentence Boundary Detection in Legal Text

Reshma Sheik, Gokul T and S Nirmala

Semantic Segmentation of Legal Documents via Rhetorical Roles

Vijit Malik, Rishabh Sanjay, Shouvik Kumar Guha, Angshuman Hazarika, Shubham Nigam, Arnab Bhattacharya and Ashutosh Modi

Detecting Relevant Differences Between Similar Legal Texts

Xiang Li, Jiaxun Gao, Diana Inkpen and Wolfgang Alschner

E-NER — An Annotated Named Entity Recognition Corpus of Legal Text

Ting Wai Terence Au, Vasileios Lampos and Ingemar Cox

12:40 - 14:00 *Lunch and In-Person Poster Session*

14:00 - 15:30 *Session 3*

On What it Means to Pay Your Fair Share: Towards Automatically Mapping Different Conceptions of Tax Justice in Legal Research Literature

Reto Gubelmann, Peter Hongler, Elina Margadant and Siegfried Handschuh

Thursday, December 8, 2022 (continued)

Combining WordNet and Word Embeddings in Data Augmentation for Legal Texts

Sezen Perçin, Andrea Galassi, Francesca Lagioia, Federico Ruggeri, Piera Santin, Giovanni Sartor and Paolo Torroni

Named Entity Recognition in Indian court judgments

Prathamesh Kalamkar, Astha Agarwal, Aman Tiwari, Smita Gupta, Saurabh Karn and Vivek Raghavan

Legal Named Entity Recognition with Multi-Task Domain Adaptation

Răzvan-alexandru Smădu, Ion-robert Dinică, Andrei-marius Avram, Dumitru-clementin Cercel, Florin Pop and Mihaela-claudia Cercel

Do Charge Prediction Models Learn Legal Theory?

Dongyan Zhao, Yansong Feng, Cong Jiang, Quzhe Huang and Zhenwei An

Legal-Tech Open Diaries: Lesson learned on how to develop and deploy light-weight models in the era of humongous Language Models

Stelios Maroudas, Sotiris Legkas, Prodromos Malakasiotis and Ilias Chalkidis

Processing Long Legal Documents with Pre-trained Transformers: Modding LegalBERT and Longformer

Dimitris Mamakas, Petros Tsotsi, Ion Androutsopoulos and Ilias Chalkidis

Privacy-Preserving Models for Legal Natural Language Processing

Ying Yin and Ivan Habernal

Automatic Classification of Legal Violations in Cookie Banner Texts

Marieke Van Hofslot, Almila Akdag Salah, Albert Gatt and Cristiana Santos

Tracking Semantic Shifts in German Court Decisions with Diachronic Word Embeddings

Daniel Braun

15:30 - 16:00 *Break*

16:00 - 17:30 *Session 4*

Should I disclose my dataset? Caveats between reproducibility and individual data rights

Raysa M. Benatti, Camila M. L. Villarroel, Sandra Avila, Esther L. Colombini and Fabiana Severi

Thursday, December 8, 2022 (continued)

Privacy Pitfalls of Online Service Terms and Conditions: a Hybrid Approach for Classification and Summarization

Emilia Lukose, Suparna De and Jon Johnson

Computing and Exploiting Document Structure to Improve Unsupervised Extractive Summarization of Legal Case Decisions

Yang Zhong and Diane Litman

Abstractive Summarization of Dutch Court Verdicts Using Sequence-to-sequence Models

Marijn Schraagen, Floris Bex, Nick Van De Luijngaarden and Daniël Prijs

Graph-based Keyword Planning for Legal Clause Generation from Topics

Sagar Joshi, Sumanth Balaji, Aparna Garimella and Vasudeva Varma

Text Simplification for Legal Domain: Insights and Challenges

Aparna Garimella, Abhilasha Sancheti, Vinay Aggarwal, Ananya Ganesh, Niyati Chhaya and Nandakishore Kambhatla

On Breadth Alone: Improving the Precision of Terminology Extraction Systems on Patent Corpora

Sean Nordquist and Adam Meyers

The Legal Argument Reasoning Task in Civil Procedure

Leonard Bongard, Lena Held and Ivan Habernal

Legal and Political Stance Detection of SCOTUS Language

Noah Bergam, Emily Allaway and Kathleen Mckeown

Can AMR Assist Legal and Logical Reasoning?

Daniel Hershcovich, Hugo López, Ruixiang Cui and Nikolaus Schrack

LawngNLI: A Long-Premise Benchmark for In-Domain Generalization from Short to Long Contexts and for Implication-Based Retrieval

Dan Roth and William Bruno

Legal Prompting: Teaching a Language Model to Think Like a Lawyer

Fangyi Yu, Lee Quartey and Frank Schilder

Thursday, December 8, 2022 (continued)

17:30 - 18:30 *From NLLP to Legal NLP - The Future of the Field (Panel) & Best Presentation Award*