EMNLP 2021

Natural Legal Language Processing (NLLP)

Proceedings of the 2021 Workshop

November 10, 2021
Punta Cana, Dominican Republic

©2021 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 978-1-954085-98-5

Introduction

Welcome to the third edition of the NLLP (Natural Legal Language Processing) Workshop, co-located with the 2021 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 We offered the option of submitting original unpublished research as non-archival in order to accommodate publication of the work at a later date in a conference or journal. These papers were reviewed following the same procedure as archival submissions. We also offered a venue for presentation for papers accepted to the Findings of EMNLP 2021 on the above topics.

We received a record number of 48 submissions and accepted 28 papers for an overall acceptance rate of 58.3% percent, all being presented orally. Out of the 28 accepted papers, 17 are long papers, 7 are short papers and 4 are original work submitted as non-archival. Each paper was reviewed by 3 or 4 members of the Program Committee. The papers cover a wide range of topics from new data sets for legal NLP and Transformer models pre-trained on legal corpora to information retrieval, extraction, question answering, classification, parsing, and summarization for legal documents to legal judgment prediction, legal dialogue, legal reasoning, and ethics.

We thank our two invited speakers for accepting our invitation: John Armour Professor of Law and Finance at Oxford University and Sylvie Delacroix, Professor in Law and Ethics at the University of Birmingham. In the tradition of past NLLP workshops, the invited speakers are legal scholars with an interest in empirical methods for legal analysis including artificial intelligence and NLP methods. We hope their talks offer a fresh perspective for the attendees.

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 2021 conference for hosting us and the workshop and publication chairs for their support. We thank our sponsors – Bloomberg and Bloomberg Law – for their contributions.

The NLLP Workshop organizers.

`http://nllpw.org`

Organizers:

Nikolaos Aletras, University of Sheffield (UK)
Ion Androutsopoulos, Athens University of Economics and Business (Greece)
Leslie Barrett, Bloomberg Law (USA)
Cătălina Goanţă, Maastricht University (The Netherlands)
Daniel Preoţiuc-Pietro, Bloomberg (USA)

Program Committee:

Tomaso Agnoloni, Institute of Legal Information Theory and Technologies (Italy)
Elliott Ash, ETH Zürich (Switzerland)
Breck Baldwin, Columbia University (US)
Andrew Blair-Stanek, Johns Hopkins University (US)
Lukasz Borchmann, Applica.ai (Poland)
Ilias Chalkidis, University of Copenhagen (Denmark)
Rajarathnam Chandramouli, Stevens Institute of Technology (US)
Daniel Chen, Toulouse School of Economics(France)
Marina Danilevsky, IBM Research (US)
Tony Davis, Southern Oregon University (US)
Stefania Degaetano-Ortlieb, Saarland University (Germany)
Luigi Di Caro, Turin University (Italy)
Fabiana Di Porto, University of Salento (Italy)
Kasper Drazewski, The European Consumer Organization (Italy)
Arthur Dyevre, Katholieke Universiteit Leuven (Belgium)
Emmanouil Fergadiotis, NCSR "Demokritos" & AUEB (Greece)
Frank Giaoui, Optimalex Legal Analytics & Columbia University (US)
Matthias Grabmair, Technical University of Münich (Germany)
Nils Holzenberger, Johns Hopkins University (US)
Dan Katz, Illinois Tech - Chicago Kent College of Law (US)
Ilan Kernerman, K Dictionaries (Israel)
Manolis Koubarakis, University of Athens (Greece)
Vasileios Lampos, University College London (UK)
Jochen Leidner, Coburg University (Germany)
Monika Leszczynska, Maastricht University (The Netherlands)
Junyi Jessy Li, University of Texas at Austin (US)
Ruta Liepina, Maastricht University (The Netherlands)
How Khang Lim, Singapore Management University (Singapore)
Prodromos Malakasiotis, NCSR "Demokritos" & AUEB (Greece)
Adam Meyers, New York University (US)
Jelena Mitrovic, University of Passau (Germany)
Paulo Quaresma, Universidade de Évora (Portugalia)
Georg Rehm, DFKI (Germany)
George Sanchez, Thomson Reuters (US)
Thibault Schrepel, Vrije Universiteit (The Netherlands)
Mathias Siems, European University Institute (Italy)
Dan Simonson, BlackBoiler LLC (US)
Jerrold Soh, Singapore Management University (Singapore)

Gerasimos Spanakis, Maastricht University (The Netherlands)
Andrea Tagarelli, University of Calabria (Italy)
Amelia Taylor, Malawi Polytechnic (Malawi)
Dimitrios Tsarapatsanis, University of York (UK)
Gijs Van Dijck, Maastricht University (The Netherlands)
Mihaela Vela, Saarland University (Germany)
Adam Wyner, Swansea University (UK)
Marcos Zampieri , Rochester Institute of Technology (US)

Invited Speakers:

John Armour, Oxford University (UK)
Sylvie Delacroix, University of Birmingham (UK)

The Power of Legislatures in Hungary - A Text Reuse Analysis

Authors: Miklós Sebők (Centre for Social Sciences, Budapest), Anna Szekély (Centre for Social Sciences, Budapest) and István Járay (Centre for Social Sciences, Budapest)

Abstract

In this paper we shed fresh light on parliaments' "viscosity": their ability to withstand government pressure when it comes to passing laws, choosing Hungary as a case of our study. We use state-of-the-art NLP methods (in the subfield of text reuse analysis) to gauge the changes of bill texts. We expect the bill's institutional and intrinsic features and partisan and coalition politics to largely influence the likelihood of bill change. Our results show that international treaties and bills submitted during electoral years are amended less likely. Furthermore, certain policy areas also decrease the likelihood of bill change, especially those which are connected to the core functions of the state: Law and order, defense, international affairs, and government operation.

LexGLUE: A Benchmark Dataset for Legal Language Understanding in English

Authors: Ilias Chalkidis (University of Copenhagen); Abhik Jana (UniversitÃ¤t Hamburg); Dirk Hartung (Bucerius Law School); Michael Bommarito (Stanford Law School); Ion Androutsopoulos (Athens University of Economics and Business); Daniel Martin Katz (Illinois Tech - Chicago Kent College of Law); Nikolaos Aletras (University of Sheffield)

Abstract

Law, interpretations of law, legal arguments, agreements, etc. are typically expressed in writing, leading to the production of vast corpora of legal text. Their analysis, which is at the center of legal practice, becomes increasingly elaborate as these collections grow in size. Natural language understanding (NLU) technologies can be a valuable tool to support legal practitioners in these endeavors. Their usefulness, however, largely depends on whether current state-of-the-art models can generalize across various tasks in the legal domain. To answer this currently open question, we introduce the Legal General Language Understanding Evaluation (LexGLUE) benchmark, a collection of datasets for evaluating model performance across a diverse set of legal NLU tasks in a standardized way. We also provide an evaluation and analysis of several generic and legal-oriented models demonstrating that the latter consistently offer performance improvements across multiple tasks.

AutoLAW: Augmented Legal Reasoning through Legal Precedent Prediction

Authors: Robert Zev Mahari (Human Dynamics Group, MIT Media Lab, Massachusetts Institute of Technology; Harvard Law School)

Abstract

This paper demonstrate how NLP can be used to address an unmet need of the legal community and increase access to justice. The paper introduces Legal Precedent Prediction (LPP), the task of predicting relevant passages from precedential court decisions given the context of a legal argument. To this end, the paper showcases a BERT model, trained on 530,000 examples of legal arguments made by U.S. federal judges, to predict relevant passages from precedential court decisions given the context of a legal argument. In 96% of unseen test examples the correct target passage is among the top-10 predicted passages. The same model is able to predict relevant precedent given a short summary of a complex and unseen legal brief, predicting the precedent that was actually cited by the brief's co-author, former U.S. Solicitor General and current U.S. Supreme Court Justice Elena Kagan.

Privacy Policy Question Answering Assistant A Query-Guided Extractive Summarization Approach

Authors: Moniba Keymanesh (The Ohio State University); Micha Elsner (The Ohio State University); Srinivasan Parthasarathy (The Ohio State University)

Abstract

Existing work on making privacy policies accessible has explored new presentation forms such as color-coding based on the risk factors or summarization to assist users with conscious agreement. To facilitate a more personalized interaction with the policies, in this work, we propose an automated privacy policy question answering assistant that extracts a summary in response to the input user query. This is a challenging task because users articulate their privacy-related questions in a very different language than the legal language of the policy, making it difficult for the system to understand their inquiry. Moreover, existing annotated data in this domain are limited. We address these problems by paraphrasing to bring the style and language of the user's question closer to the language of privacy policies. Our content scoring module uses the existing in-domain data to find relevant information in the policy and incorporates it in a summary. Our pipeline is able to find an answer for 89% of the user queries in the privacyQA dataset.

Table of Contents

A Corpus for Multilingual Analysis of Online Terms of Service
Kasper Drawzeski, Andrea Galassi, Agnieszka Jablonowska, Francesca Lagioia, Marco Lippi, Hans Wolfgang Micklitz, Giovanni Sartor, Giacomo Tagiuri and Paolo Torroni 1

Named Entity Recognition in the Romanian Legal Domain
Vasile Pais, Maria Mitrofan, Carol Luca Gasan, Vlad Coneschi and Alexandru Ianov 9

Swiss-Judgment-Prediction: A Multilingual Legal Judgment Prediction Benchmark
Joel Niklaus, Ilias Chalkidis and Matthias Stürmer 19

Automated Extraction of Sentencing Decisions from Court Cases in the Hebrew Language
Mohr Wenger, Tom Kalir, Noga Berger, Carmit Klar Chalamish, Renana Keydar and Gabriel Stanovsky 36

A Multilingual Approach to Identify and Classify Exceptional Measures against COVID-19
Georgios Tziafas, Eugenie de Saint-Phalle, Wietse de Vries, Clara Egger and Tommaso Caselli . 46

Multi-granular Legal Topic Classification on Greek Legislation
Christos Papaloukas, Ilias Chalkidis, Konstantinos Athinaios, Despina Pantazi and Manolis Koubarakis 63

Machine Extraction of Tax Laws from Legislative Texts
Elliott Ash, Malka Guillot and Luyang Han 76

jurBERT: A Romanian BERT Model for Legal Judgement Prediction
Mihai Masala, Radu Cristian Alexandru Iacob, Ana Sabina Uban, Marina Cidota, Horia Velicu, Traian Rebedea and Marius Popescu 86

JuriBERT: A Masked-Language Model Adaptation for French Legal Text
Stella Douka, Hadi Abdine, Michalis Vazirgiannis, Rajaa El Hamdani and David Restrepo Amariles 95

Few-shot and Zero-shot Approaches to Legal Text Classification: A Case Study in the Financial Sector
Rajdeep Sarkar, Atul Kr. Ojha, Jay Megaro, John Mariano, Vall Herard and John P. McCrae . . 102

A Free Format Legal Question Answering System
Soha Khazaeli, Janardhana Punuru, Chad Morris, Sanjay Sharma, Bert Staub, Michael Cole, Sunny Chiu-Webster and Dhruv Sakalley 107

Searching for Legal Documents at Paragraph Level: Automating Label Generation and Use of an Extended Attention Mask for Boosting Neural Models of Semantic Similarity
Li Tang and Simon Clematide 114

GerDaLIR: A German Dataset for Legal Information Retrieval
Marco Wrzalik and Dirk Krechel 123

SPaR.txt, a Cheap Shallow Parsing Approach for Regulatory Texts
Ruben Kruiper, Ioannis Konstas, Alasdair J.G. Gray, Farhad Sadeghineko, Richard Watson and Bimal Kumar 129

Capturing Logical Structure of Visually Structured Documents with Multimodal Transition Parser
Yuta Koreeda and Christopher Manning 144

Legal Terminology Extraction with the Termolator
Nhi Pham, Lachlan Pham and Adam L. Meyers 155

Supervised Identification of Participant Slots in Contracts
Dan Simonson 163

Named Entity Recognition in Historic Legal Text: A Transformer and State Machine Ensemble Method
Fernando Trias, Hongming Wang, Sylvain Jaume and Stratos Idreos 172

Summarization of German Court Rulings
Ingo Glaser, Sebastian Moser and Florian Matthes 180

Learning from Limited Labels for Long Legal Dialogue
Jenny Hong, Derek Chong and Christopher Manning 190

Automating Claim Construction in Patent Applications: The CMUmine Dataset
Ozan Tonguz, Yiwei Qin, Yimeng Gu and Hyun Hannah Moon 205

Effectively Leveraging BERT for Legal Document Classification
Nut Limsopatham 210

Semi-automatic Triage of Requests for Free Legal Assistance
Meladel Mistica, Jey Han Lau, Brayden Merrifield, Kate Fazio and Timothy Baldwin 217

Automatic Resolution of Domain Name Disputes
Wayan Oger Vihikan, Meladel Mistica, Inbar Levy, Andrew Christie and Timothy Baldwin . . . 228

Conference Program

08:00–08:10 *Workshop Opening*

08:10–09:00 *Invited Talk*
John Armour

09:00–09:15 *A Corpus for Multilingual Analysis of Online Terms of Service*
Kasper Drawzeski, Andrea Galassi, Agnieszka Jablonowska, Francesca Lagioia, Marco Lippi, Hans Wolfgang Micklitz, Giovanni Sartor, Giacomo Tagiuri and Paolo Torroni

09:15–09:30 *Named Entity Recognition in the Romanian Legal Domain*
Vasile Pais, Maria Mitrofan, Carol Luca Gasan, Vlad Coneschi and Alexandru Ianov

09:30–09:45 *The Power of Legislatures in Hungary - A Text Reuse Analysis*
Miklós Sebők, Anna Szekély and István Járay

09:45–10:00 *Swiss-Judgment-Prediction: A Multilingual Legal Judgment Prediction Benchmark*
Joel Niklaus, Ilias Chalkidis and Matthias Stürmer

10:00–10:15 *LexGLUE: A Benchmark Dataset for Legal Language Understanding in English*
Ilias Chalkidis, Abhik Jana, Dirk Hartung, Michael Bommarito, Ion Androutsopoulos, Daniel Martin Katz and Nikolaos Aletras

10:15–10:30 *'Just What do You Think You're Doing, Dave?' A Checklist for Responsible Data Use in NLP*
Anna Rogers, Tim Baldwin and Kobi Leins

10:30–10:45 *Break*

10:45–11:00 *Automated Extraction of Sentencing Decisions from Court Cases in the Hebrew Language*
Mohr Wenger, Tom Kalir, Noga Berger, Carmit Klar Chalamish, Renana Keydar and Gabriel Stanovsky

11:00–11:15 *A Multilingual Approach to Identify and Classify Exceptional Measures against COVID-19*
Georgios Tziafas, Eugenie de Saint-Phalle, Wietse de Vries, Clara Egger and Tommaso Caselli

11:15–11:30 *Multi-granular Legal Topic Classification on Greek Legislation*
Christos Papaloukas, Ilias Chalkidis, Konstantinos Athinaios, Despina Pantazi and Manolis Koubarakis

11:30–11:40 *Machine Extraction of Tax Laws from Legislative Texts*
Elliott Ash, Malka Guillot and Luyang Han

10 November 2021

11:40–11:50 *jurBERT: A Romanian BERT Model for Legal Judgement Prediction*
Mihai Masala, Radu Cristian Alexandru Iacob, Ana Sabina Uban, Marina Cidota, Horia Velicu, Traian Rebedea and Marius Popescu

11:50–12:00 *JuriBERT: A Masked-Language Model Adaptation for French Legal Text*
Stella Douka, Hadi Abdine, Michalis Vazirgiannis, Rajaa El Hamdani and David Restrepo Amariles

12:00–12:10 *Few-shot and Zero-shot Approaches to Legal Text Classification: A Case Study in the Financial Sector*
Rajdeep Sarkar, Atul Kr. Ojha, Jay Megaro, John Mariano, Vall Herard and John P. McCrae

12:10–13:00 *Lunch Break & Virtual Town Hall*

13:00–13:45 *Invited talk: Data Trusts as a Bottom-up Empowerment Tool*
Sylvie Delacroix

13:45–14:00 *AutoLAW: Augmented Legal Reasoning through Legal Precedent Prediction*
Robert Zev Mahari

14:00–14:10 *A Free Format Legal Question Answering System*
Soha Khazaeli, Janardhana Punuru, Chad Morris, Sanjay Sharma, Bert Staub, Michael Cole, Sunny Chiu-Webster and Dhruv Sakalley

14:10–14:20 *Searching for Legal Documents at Paragraph Level: Automating Label Generation and Use of an Extended Attention Mask for Boosting Neural Models of Semantic Similarity*
Li Tang and Simon Clematide

14:20–14:30 *GerDaLIR: A German Dataset for Legal Information Retrieval*
Marco Wrzalik and Dirk Krechel

14:30–14:45 *Break*

14:45–15:00 *SPaR.txt, a Cheap Shallow Parsing Approach for Regulatory Texts*
Ruben Kruiper, Ioannis Konstas, Alasdair J.G. Gray, Farhad Sadeghineko, Richard Watson and Bimal Kumar

15:00–15:15 *Capturing Logical Structure of Visually Structured Documents with Multimodal Transition Parser*
Yuta Koreeda and Christopher Manning

15:15–15:30 *Legal Terminology Extraction with the Termolator*
Nhi Pham, Lachlan Pham and Adam L. Meyers

10 November 2021

15:30–15:45 *Supervised Identification of Participant Slots in Contracts*
Dan Simonson

15:45–16:00 *ContractNLI: A Dataset for Document-level Natural Language Inference for Contracts*
Yuta Koreeda and Christopher Manning

16:00–16:10 *Named Entity Recognition in Historic Legal Text: A Transformer and State Machine Ensemble Method*
Fernando Trias, Hongming Wang, Sylvain Jaume and Stratos Idreos

16:10–16:25 *Break*

16:25–16:40 *Summarization of German Court Rulings*
Ingo Glaser, Sebastian Moser and Florian Matthes

16:40–16:55 *Privacy Policy Question Answering Assistant A Query-Guided Extractive Summarization Approach*
Moniba Keymanesh, Micha Elsner and Srinivasan Parthasarathy

16:55–17:10 *Learning from Limited Labels for Long Legal Dialogue*
Jenny Hong, Derek Chong and Christopher Manning

17:10–17:20 *Automating Claim Construction in Patent Applications: The CMUmine Dataset*
Ozan Tonguz, Yiwei Qin, Yimeng Gu and Hyun Hannah Moon

17:20–17:30 *Effectively Leveraging BERT for Legal Document Classification*
Nut Limsopatham

17:30–17:45 *Semi-automatic Triage of Requests for Free Legal Assistance*
Meladel Mistica, Jey Han Lau, Brayden Merrifield, Kate Fazio and Timothy Baldwin

17:45–18:00 *Automatic Resolution of Domain Name Disputes*
Wayan Oger Vihikan, Meladel Mistica, Inbar Levy, Andrew Christie and Timothy Baldwin