NAACL HLT 2019

Natural Legal Language Processing (NLLP)

Proceedings of the 2019 Workshop

June 7, 2019
Minneapolis, MN

©2019 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 978-1-950737-03-1

Introduction

Welcome to the first edition of the NLLP (Natural Legal Language Processing) Workshop, co-located with the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

Many industries have embraced natural language processing (NLP) approaches, which have altered healthcare, finance, education and other fields. The legal domain however remains largely underrepresented in the NLP literature despite its 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 documents. 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 natural language processing and legal backgrounds.

We were interested in five 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. We also offered the option of submitting original unpublished research as non-archival in order to accommodate publication of the work at a later date in a conference or journal.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 received 20 submissions and accepted 12 papers for an overall acceptance rate of 60 percent, all being presented orally. Out of the 12 accepted papers, 6 are long papers, 3 are short papers and 3 are original work submitted as non-archival. Half of the accepted papers have primarily industry authors. Each paper was reviewed by 3 to 5 members of the Program Committee. The papers cover a range of topics including bias in the judiciary, predictive methods for legal documents, building NLP tools to process legal documents and system descriptions for processing contracts or dockets.

We thank our two invited speakers for accepting our invitation. Both speakers are legal scholars with an interest in using artificial intelligence and natural language processing methods for legal analysis. We hope their talks offer a fresh perspective for the attendees. Prof. Arthur Dyevre presents a talk titled: 'Law as Data: The Promise and Challenges of Natural Language Processing for Legal Research' and Prof. Daniel M. Katz presents a talk titled: 'NLP & Law {Past, Present + Future}'.

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, NAACL HLT 2019 for hosting us and the local, workshop and publication chairs for their support. We especially thank our sponsors – Bloomberg and Bloomberg Law – for their contributions.

We are looking forward to meeting the authors and the other participants in the workshop in Minneapolis.

The NLLP Workshop organizers.

`http://nllpw.org`

Organizers:

Nikolaos Aletras, University of Sheffield (UK)
Elliott Ash, ETH Zurich (Switzerland)
Leslie Barrett, Bloomberg Law (USA)
Daniel L. Chen, Toulouse School of Economics (France)
Adam Meyers, New York University (USA)
Daniel Preoţiuc-Pietro, Bloomberg LP (USA)
David Rosenberg, Bloomberg LP (USA)
Amanda Stent, Bloomberg LP (USA)

Program Committee:

Tomaso Agnoloni, Institute of Legal Information Theory and Technologies (Italy)
Ion Androutsopoulos, Athens University of Economics and Business (Greece)
Joan Bachenko, Linguistech LLC (USA)
Claire Cardie, Cornell University (USA)
Ilias Chalkidis, Athens University of Economics and Business (Greece)
Rajarathnam Chandramouli, Stevens Institute of Technology (USA)
Laura Chiticariu, IBM Research (USA)
Walter Daelemans, University of Antwerp (Belgium)
Marina Danilevsky, IBM Research (USA)
Stefania Degaetano-Ortlieb, Saarland University (Germany)
Luigi Di Caro, University of Turin (Italy)
Liviu P. Dinu, University of Bucharest (Romania)
Eileen Fitzpatrick, Montclair State University (USA)
Enrico Francesconi, Institute of Legal Information Theory and Technologies (Italy)
Frank S. Giaoui, Columbia Law School (USA)
Matthias Grabmair, Carnegie Mellon University (USA)
Ilan Kernerman, K Dictionaries (Israel)
Seth Kulick, University of Pennsylvania (USA)
Vasileios Lampos, University College London (UK)
Mark Liberman, University of Pennsylvania (USA)
Shervin Malmasi, Harvard Medical School (USA)
Elena Montiel, Universidad Politécnica de Madrid (Spain)
Paulo Quaresma, University of Évora (Portugal)
Georg Rehm, DFKI (Germany)
Victor Rodríguez-Doncel, Universidad Politécnica de Madrid (Spain)
Victoria Rubin, University of Western Ontario (Canada)
Eugene Santos, Thayer School of Engineering at Dartmouth (USA)
Maosong Sun, Tsinghua University (China)
Dimitrios Tsarapatsanis, University of York (UK)
Mihaela Vela, Saarland University (Germany)
Marc B. Vilain, MITRE Corp (USA)
Jianqian Wang, SUNY Buffalo (USA)
Adam Wyner, Swansea University (UK)
Diyi Yang, Georgia Institute of Technology (USA)
Marcos Zampieri, University of Wolverhampton (UK)

Invited Speakers:

Arthur Dyevre, Katholieke Universiteit Leuven (Belgium)
Daniel Martin Katz, Illinois Institute of Technology – Chicago Kent College of Law (USA)

Table of Contents

Plain English Summarization of Contracts
Laura Manor and Junyi Jessy Li 1

Scalable Methods for Annotating Legal-Decision Corpora
Lisa Ferro, John Aberdeen, Karl Branting, Craig Pfeifer, Alexander Yeh and Amartya Chakraborty 12

The Extent of Repetition in Contract Language
Dan Simonson, Daniel Broderick and Jonathan Herr 21

Sentence Boundary Detection in Legal Text
George Sanchez 31

Legal Linking: Citation Resolution and Suggestion in Constitutional Law
Robert Shaffer and Stephen Mayhew 39

Litigation Analytics: Case Outcomes Extracted from US Federal Court Dockets
Thomas Vacek, Ronald Teo, Dezhao Song, Timothy Nugent, Conner Cowling and Frank Schilder 45

Developing and Orchestrating a Portfolio of Natural Legal Language Processing and Document Curation Services
Georg Rehm, Julian Moreno-Schneider, Jorge Gracia, Artem Revenko, Victor Mireles, Maria Khvalchik, Ilan Kernerman, Andis Lagzdins, Marcis Pinnis, Artus Vasilevskis, Elena Leitner, Jan Milde and Pia Weißenhorn 55

Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments
Jerrold Soh, How Khang Lim and Ian Ernst Chai 67

Extreme Multi-Label Legal Text Classification: A Case Study in EU Legislation
Ilias Chalkidis, Emmanouil Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras and Ion Androutsopoulos 78

Conference Program

Friday, June 7, 2019

9:00–9:10 ***Workshop Opening***

9:10–10:10 ***Invited Speaker***

9:10–10:10 *Law as Data: The Promise and Challenges of Natural Language Processing for Legal Research*
Arthur Dyevre

10:10–10:30 ***Bias***

10:10–10:30 *Implicit Bias in the Judiciary*
Daniel Chen, Elliott Ash and Arianna Ornaghi

10:30–11:00 ***Coffee***

11:00–12:20 ***NLP Applications***

11:00–11:20 *Plain English Summarization of Contracts*
Laura Manor and Junyi Jessy Li

11:20–11:40 *Question Answering for Privacy Policies: Combining Computational and Legal Perspectives*
Abhilasha Ravichander, Alan W Black, Thomas Norton, Shomir Wilson and Norman Sadeh

11:40–12:00 *Scalable Methods for Annotating Legal-Decision Corpora*
Lisa Ferro, John Aberdeen, Karl Branting, Craig Pfeifer, Alexander Yeh and Amartya Chakraborty

12:00–12:20 *The Extent of Repetition in Contract Language*
Dan Simonson, Daniel Broderick and Jonathan Herr

12:20–14:00 ***Lunch***

Friday, June 7, 2019 (continued)

14:00–15:00 ***Invited Speaker***

14:00–15:00 *NLP & Law {Past, Present + Future}*
Daniel Katz

15:00–15:30 ***Short Papers***

15:00–15:15 *Sentence Boundary Detection in Legal Text*
George Sanchez

15:15–15:30 *Legal Linking: Citation Resolution and Suggestion in Constitutional Law*
Robert Shaffer and Stephen Mayhew

15:30–16:00 ***Coffee***

16:00–17:00 ***Demos***

16:00–16:20 *Litigation Analytics: Case Outcomes Extracted from US Federal Court Dockets*
Thomas Vacek, Ronald Teo, Dezhao Song, Timothy Nugent, Conner Cowling and Frank Schilder

16:20–16:40 *Developing and Orchestrating a Portfolio of Natural Legal Language Processing and Document Curation Services*
Georg Rehm, Julian Moreno-Schneider, Jorge Gracia, Artem Revenko, Victor Mireles, Maria Khvalchik, Ilan Kernerman, Andis Lagzdins, Marcis Pinnis, Artus Vasilevskis, Elena Leitner, Jan Milde and Pia Weißenhorn

16:40–17:00 *Transparent Linguistic Models for Contract Understanding and Comparison*
Arvind Agarwal, Laura Chiticariu, Poornima Chozhiyath Raman, Marina Danilevsky, Diman Ghazi, Ankush Gupta, Shanmukh Guttula, Yannis Katsis, Rajasekar Krishnamurthy, Yunyao Li, Shubham Mudgal, Vitobha Munigala, Nicholas Phan, Dhaval Sonawane, Sneha Srinivasan, Sudarshan Thitte, Shivakumar Vaithyanathan, Mitesh Vasa, Ramiya Venkatachalam, Vinitha Yaski and Huaiyu Zhu

17:00–17:10 ***Break***

17:10–17:50 ***Prediction***

17:10–17:30 *Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments*
Jerrold Soh, How Khang Lim and Ian Ernst Chai

Friday, June 7, 2019 (continued)

17:30–17:50 *Extreme Multi-Label Legal Text Classification: A Case Study in EU Legislation*
Ilias Chalkidis, Emmanouil Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras and Ion Androutsopoulos

17:50–18:00 ***Workshop Closing***