International Workshop on Legal Data Analytics and Mining (LeDAM 2018): Preface to the Proceedings

Arindam Pal*   Arnab Bhattacharya§   Indrajit Bhattacharya*

Kripabandhu Ghosh§   Lipika Dey*   Marie-Francine Moens†

Saptarshi Ghosh‡

Tata Consultancy Services Research, India*
Indian Institute of Technology Kanpur, India§
Katholieke Universiteit Leuven, Belgium†
Indian Institute of Technology Kharagpur, India‡

1 INTRODUCTION

Legal data mining is the subarea of data mining applied to legal texts, such as legislation, case law, patents, and scholarly works. Legal data mining systems are important to provide easier access to and insights about law for both common persons and legal professionals. This area is becoming increasingly important, because of the rapidly growing volume of legal cases and documents available in digital formats. For this reason, we organized the First International Workshop on Legal Data Analytics and Mining (LeDAM 2018), co-located with ACM CIKM 2018. The website of LeDAM 2018 is https://sites.google.com/site/legaldam2018/. The objectives of the LeDAM 2018 workshop are to: (1) Provide a venue for academic and industrial/governmental researchers and professionals to come together, present and discuss research results, use cases, innovative ideas, challenges, and opportunities that arise from applications of data mining in the legal domain, and (2) Foster collaborations between the Legal and the Artificial Intelligence, Data Mining, Information Retrieval, and Machine Learning communities.

The workshop programme included invited talks by the following reputed researchers (see Section 2 for details):

- Giovanni Sartor, Professor of Legal Informatics and Legal Theory, European University Institute, Italy
- Luigi Di Caro, Assistant Professor, Department of Computer Science, University of Turin, Italy
- Jack G. Conrad, Lead Research Scientist, Center for AI and Cognitive Computing, Thomson Reuters Labs, USA

The program also included presentation of papers accepted through the peer-reviewed track (see Section 3), and a panel discussion on emerging problems in legal data mining. We specifically attempted to ensure the presence of both academicians from the data mining/IR/ML communities as well as practitioners from the Law industry among our invited speakers and members of our Program Committee (stated in Section 3). For further details, refer to the LeDAM 2018 website https://sites.google.com/site/legaldam2018/.

2 DETAILS OF INVITED TALKS

The LeDAM 2018 workshop included the following invited talks.

- **Speaker:** Giovanni Sartor, Professor of Legal Informatics and Legal Theory, European University Institute, Italy
  Title: Using Machine Learning to Support Law Enforcement to the Benefit of Consumers and Data Subject: the CLAUDETTE Project
  Abstract: The project CLAUDETTE aims to support the detection of potentially unfair and unlawful clause, both in consumer contacts and in privacy policies, through automated tools, based on computational linguistic and artificial intelligence. The purpose is to enable consumer protection bodies and data protection authorities to engage more proactively and effectively in monitoring compliance and in enforcing the law. With regard to both contract terms and privacy policy we have collected a corpus of contract terms, identified different kinds of unlawful and unfair terms through legal analysis, and annotated the documents accordingly. Then we have applied and tested different computational approaches, including various machine learning algorithms, to detect such terms. The better performing algorithms have been implemented in an application available to the public through the project's web site. The system is complemented by a crawler, that detects changes in the contract and policies already submitted to the system.

- **Speaker:** Luigi Di Caro, Assistant Professor, Department of Computer Science, University of Turin, Italy
  Title: Natural Language Processing and Ontology Learning in the Legal Domain
  Abstract: Legal ontologies aim to provide a structured representation of legal concepts and their interconnections. These ontologies are then exploited to support tasks such as information extraction and question answering in the legal domain. Given the increasing importance of the Web of Data in public administration and in companies, being able to provide machine-readable legal information is becoming a

valuable and desired contribution. However, concepts and relations within existing ontologies usually represent limited subjective and application-oriented views of specific subdomains of interest. The talk will discuss resent research on natural language technologies and text mining approaches towards the creation, the reuse and the enrichment of legal ontologies.

- **Speaker:** Jack G. Conrad, Lead Research Scientist, Center for AI and Cognitive Computing, Thomson Reuters, USA
  Title: 30 Years of AI and Law: Legal Data Analytics in the Long View – Looking Back, Looking Forward
  Abstract: This talk will begin by examining the roots of Artificial Intelligence and Law – including applications involving NLP, data mining, machine learning, and more broadly, data analytics – noting that it has been around for much longer than the recent buzz would suggest. We will explore the field of AI and Law in terms of its development and expansion starting in the 1980s and study how seminal research was conducted and reported on in conference proceedings such as ICAIL and publications such as the AI and Law journal. After having established the foundations of today's field of AI and Law, we will look to the future and sketch some of the practical application scenarios that the capabilities from the field promise to deliver. These include next-generation tools for legal professionals that can augment their skill sets by providing analytical abilities to help in the crafting of legal strategies. We will illustrate such instruments through the visualization of expected outcomes, while varying key parameters such as trial length, expected costs, and likely award or settlement figures. Lastly, we will investigate the prospective role that prediction tools can play in AI and Law application spaces, while looking still further into the future.

3 PEER-REVIEWED PAPER TRACK

Eight papers were submitted to the peer-review track, from diverse countries all over the world. Each submitted paper was reviewed by at least three members of the following Program Committee:

- Adam Wyner, Swansea University, Swansea, UK
- Charles K. Nicholas, University of Maryland Baltimore County, USA
- Dave Lewis, Brainspace - A Cyxtera Business, USA
- Girish Keshav Palshikar, Tata Consultancy Services, India
- Haozhen Zhao, Legal Technology Solution Practice, Navigant
- Jack G. Conrad, Thomson Reuters, USA
- Jeroen Keppens, King's College London, UK

- Karl Branting, MITRE Corporation, USA
- Katie Atkinson, University of Liverpool, UK
- Ken Satoh, National Institute of Informatics, Japan
- Kevin Ashley, University of Pittsburgh, USA
- Matthias Grabmair, Carnegie Mellon University, USA
- Maura Grossman, University of Waterloo, Canada
- Mi-Young Kim, University of Alberta, Canada
- Mossab Bagdouri, Walmart Labs, USA
- Paulo Quaresma, Universidade de Evora, Portugal
- Prasenjit Majumder, DAIICT, India
- William Webber, William Webber Consulting, Australia

Five papers were accepted through the peer-review process. The papers were on various topics, including contract renewals, concept hierarchy extraction, patent clustering, argumentation-driven information extraction, deep ensemble learning. The list of papers accepted in LeDAM 2018 is as follows.

- **Title:** Structural Analysis of Contract Renewals
  Authors: Frieda Josi and Christian Wartena

- **Title:** Concept Hierarchy Extraction from Legal Literature
  Authors: Sabine Wehnert, David Broneske, Stefan Langer and Gunter Saake

- **Title:** Use of Pseudo Relevance Feedback for Patent Clustering with Fuzzy C-means
  Authors: Noushin Fadaei and Thomas Mandl

- **Title:** Argumentation-driven information extraction for online crime reports
  Authors: Marijn Schraagen, Bas Testerink, Daphne Odekerken and Floris Bex

- **Title:** Deep Ensemble Learning for Legal Query Understanding
  Authors: Arunprasath Shankar and Venkata Nagaraju Buddarapu

4 ACKNOWLEDGEMENTS