Introduction for computer law and security review: special issue “knowledge management for law”

(Article begins on next page)

Introduction for computer law and security review: special issue "knowledge management for law"

Emilio Sulis[a], Luigi Di Caro[a], Rohan Nanda[b]

[a] *Computer Science Department, University of Torino, Torino, Italy*
[b] *Maastricht Law and Tech Lab, Maastricht University, Maastricht, The Netherlands*

In today's rapidly evolving legal landscape, the intersection of Knowledge Management and Artificial Intelligence methods is becoming increasingly important for a multitude of sub-domains within legal informatics. These areas encompass Natural Language Processing (NLP), ontologies, argumentation, and more. The marriage of advanced technologies and knowledge management principles holds the promise of revolutionizing how legal knowledge is harnessed, represented, and applied. This special issue, titled *Knowledge Management for Law (KM4LAW)*, is dedicated to exploring the multifaceted realm of legal informatics through the lenses of Artificial Intelligence, Knowledge Modelling, and Information Extraction.

The topics addressed in this special issue underline the open challenges in the field, which involve researchers and practitioners in a common effort to seek models and solutions that reconcile the most advanced computational techniques with practical needs.

This KM4LAW special issue addresses the above mentioned wide range of approaches applied to legal texts. We received a total of 15 papers, a good number indicating that the topic is relevant and the subject of research. Finally, 11 articles passed the review process.

The articles that have been selected collectively contribute to the growing body of knowledge management practices in law, bringing forth cutting-edge research and practical applications in the field. Spanning a diverse array of topics, these contributions showcase the transformative potential of NLP, ontological modeling, argumentation analysis, and more, in the context of legal informatics.

Email addresses: `emilio.sulis@unito.it` (Emilio Sulis), `luigi.dicaro@unito.it` (Luigi Di Caro), `r.nanda@maastrichtuniversity.nl` (Rohan Nanda)

With the above in mind, we briefly summarize the following articles and their contribution to the topics of focus in this special issue of legal informatics:

Pairing EU directives and their national implementing measures: A dataset for semantic search

by Roger Ferrod, Denys Amore Bondarenko, Davide Audrito, Giovanni Siragusa.

The paper presents a high-quality dataset very meaningful in legal domain large corpora. The dataset contains meaningful textual data in multiple languages. The dataset includes European Directives (EUDs) and their National Implementing Measures (NIMs) for text retrieval applications. The authors demonstrated the effectiveness of the dataset through experiments with baselines.

Improving colloquial case legal judgment prediction via abstractive text summarization

by Yu-Xiang Hong, Chia-Hui Chang.

The paper describes how abstractive text summarization can be used for the task of legal judgement prediction. In the current use of AI in the legal domain, the paper addresses the issue of the reliance on the legal terminology, that might be unclear to non-professionals. The models developed in the paper expressed an idea fairly innovative, with a limited number of research sharing the same goal.

Fine-tuning GPT-3 for legal rule classification

by Davide Liga, Livio Robaldo.

The paper describes an innovative approach to the problem of classifying legal rules using the GPT-3 language model. The authors demonstrate the effectiveness of their proposed approach by training and testing the Legal Rule Classification task on the GDPR encoded in two widely used XML standards for the legal domain, i.e. LegalDocML and LegalRuleML. The paper shows that GPT-3 is capable of outperforming previous experiments on the same task and can easily learn to classify legal and deontic rules even on small amounts of data.

Public tenders, complaints, machine learning and recommender systems: a case study in public administration

by Roberto Nai, Rosa Meo, Gabriele Morina, Paolo Pasteris.

The paper presents a legal prediction and recommendation experiment over an Italian legal dataset. In particular, a machine learning-based application for e-procurement systems in the public sector. Authors explored

both a classification and recommendation system by using recent techniques like BERT for semantic text embeddings.

Enforcing legal information extraction through context-aware techniques: the ASKE Approach

by Silvana Castano, Alfio Ferrara, Emanuela Furiosi, Stefano Montanelli, Sergio Picascia, Davide Riva, Carolina Stefanetti.

The paper proposes an approach to extract relevant terminology in the legal domain used to annotate documents. The unsupervised multi-stage approach uses a thesaurus and Large Language Models (LLMs) by adopting both similarity and clustering techniques. The iterative three-step framework progressively extends and refines the terminology also by constructing relationships between terms. The experiments outperforms BERTopic on an EurLex dataset of 45k legislative documents annotated with labels from the EuroVoc thesaurus.

Legal knowledge management for prosecutors based on judgement predication and error analysis from indictments

by Kuo-Chun Chien, Chia-Hui Chang, Ren-Der Sun.

The paper presents a novel contribution in the form of a new indictment dataset for Legal Judgement Prediction in the Taiwanese juridical system. Authors demonstrate how such dataset improve the performance of a legal judgement prediction task for assisting prosecutors, as opposed to commonly used verdict documents. The experiments include two models: a multi-task learning model that uses message passing to capture dependencies between different tasks, and a prompt-based model that aims to improve performance on under-represented charges. An extensive error analysis highlights the limitations of machine learning approaches, shedding light on potential areas of improvement.

An entity-centric approach to manage court judgements based on Natural Language Processing

by Valerio Bellandi, Christian Bernasconi, Fausto Lodi, Matteo Palmonari, Riccardo Pozzi, Marco Ripamonti, Stefano Siccardi.

This paper introduces an entity-centric infrastructure for the management of legal documents, such as court judgements. It exploits a repository of textual documents and annotates them using a combination of NLP services, including machine learning and syntactic rules. The infrastructure is designed for sustainability and collaboration, involving scientific teams, institutional bodies and companies in its co-design. The paper presents experiments demonstrating the viability of the system, particularly in annotating

data with low-resource methods. It highlights a human-in-the-loop approach to align with the Italian court processes and discusses the scalability of and robustness challenges for the proposed solution.

Data modelling as a means of power: At the legal and computer science crossroads

by Kai von Lewinski, Michael Beurskens, Stefanie Scherzinger.

The paper presents a research work on conceptual data modelling and its impact on the real meaning of stored data. Authors assess how the technical framework adopted for storing, modelling, and accessing legal knowledge in digital format contributes to "shaping the state of the real world", i.e. influencing the real meaning of stored data. The technical aspects have to be accounted for by law as they have an impact on the real world. The authors provide several examples (mostly using German public registers) in which non-adequate modelling generated severe negative consequences which required a significant investment of time and efforts to resolve.

LegalHTML: Semantic mark-up of legal acts using web technologies

by Armando Stellato, Manuel Fiorelli.

The paper presents a new mark-up language for representing legal acts, i.e. an extension of the HTML language (LegalHTML). The author provides an overview of the proposed tagging language, including an ontology, the specific syntax of the language, and some usage examples. The method offers an interesting solution in terms of structure, semantics and representation.

ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the Italian legal domain

by Daniele Licari, Giovanni Comandè.

This paper describes the process to collect and select data to train a large language model of Italian legal language. The main contribution is a small collection of language models created by pre-training BERT on data collected from Italian courts. Authors included a distinction with respect of the type of law they model (civil vs. criminal) and also a distilled model. Finally, the models are tested on a series of NLP tasks in the legal domain, e.g. legal named entity recognition, or the identification of legal topics in documents. In all tests, the novel models show to perform better than the baseline, a general-purpose Italian BERT model.

An ontological-driven models for representing image-based sexual abuses

by Mattia Falduti, Christine Griffo.

This research aims to enhance the ontological analysis, modeling, and representation of image-based sexual abuses. The effort recognize the transnational and global nature of online abuses and the consequent absence of effective common legislation, whereas some countries already introduced dedicated regulation. A set of machine-readable models based on the analysis of several pieces of legislation and legal principles can be used by digital platforms as a means of monitoring and managing the compliance of users' conduct. Authors use the ontological approach to model some types of abusive conducts, building on the Unified Foundational Ontology - UFO and a core ontology of legal relations called UFO-L. The topic of the paper is scientifically relevant from an ontological theoretical perspective. Moreover, it provide insights to providers of online intermediary services who need to moderate user-generated online content.