Examining the Technical, Legal and Ethical Implications of Improved Access to Legal Information Using Supercomputing Technology: The ManyLaws Project

Shefali VIRKAR[a,1] Anna-Sophie NOVAK[b,2] Charalampos ALEXOPOULOS[c] Yannis CHARALABIDIS[d] and Michalis LOUTSARIS[e]

[a] *Danube University Krems, Austria*
[b] *Danube University Krems, Austria*
[c] *University of the Aegean, Greece*
[d] *University of the Aegean, Greece*
[e] *University of the Aegean, Greece*

Abstract. Legal information is an asset for decision making not only by EU institutions but also by member states, local administrations, businesses and citizens. Accurate, target-orientated, and timely information could enhance the digitisation of decision-making processes. Data such as legislation acts, bills, case laws, resolutions and decisions, published in each Member States' language, as well as administration and citizen-generated content is increasingly embedded in large amounts of textual data available on the Internet. The vision of ManyLaws is to produce semantically annotated Big Legal Open Data, easily searchable and exploitable based on text mining tools and algorithms offered through proper visualization techniques. The proposed workshop will focus on the factors that can contribute to the effective delivery of the new services, together with the legal and ethical implications associated with the application of advanced computing technologies to the acquisition, storage, and processing of legal information.

Keywords. Big Legal Open Data, Text Mining, Supercomputing, Ethical and Legal Implications

1. Motivation for the Workshop

Although society is overwhelmed with an overload of legal information, only legal experts can follow the latest legislation and case law produced by parliaments and courts on a national and on a European level. Accurate, target-orientated, and timely information is needed not only by EU institutions but also by member states, local

---

[1] Shefali Virkar, Department for E-Governance and Administration, Danube University Krems, Dr.-Karl-Dorrek-Straße 30, 3500 Krems an der Donau, Austria, E-mail: shefali.virkar@donau-uni.ac.at.

[2] Anna-Sophie Novak, Department for E-Governance and Administration, Danube University Krems, Dr.-Karl-Dorrek-Straße 30, 3500 Krems an der Donau, Austria, E-Mail: anna-sophie.novak@donau.uni.ac.at.

administrations, businesses and citizens at almost every stage of the decision making process.

Such, by policy makers required and/or produced data, is increasingly embedded in large amounts of textual data available on the Internet. Furthermore, the large amount of information concerning laws that apply in the EU countries currently remains fragmented across multiple national databases or inaccessible systems, mainly consisting of documents (legislation acts, bills, case laws, resolutions, decisions) published in each Member States' language. In addition, administration-generated content (e.g. local communications, regulations), citizen-generated relevant content (e.g. blogs, newsletters, social media posts) and news published in EU member states concerning legal events (e.g. law publication, draft law deliberation, EU directive publication) could be considered of major importance in every-day or in mid- and long-term decision making. It is estimated the above database will contain more than 1 trillion words in 21 different languages, corresponding to about 10 million "volumes" of classical books, when another 5,000 such "volumes" will be added for study on a daily basis.

Due to the sheer volume of data, the manual extraction of the relevant data it contains is nearly impossible. Text mining and analysis tools become necessary to address the problem of volume, of currentness, and in order to provide the right information in the proper format. The information processing stage of such an infrastructure should make use of massively parallel computing tools, balancing the load between batch and real-time service modes. Two stages should be supported: (i) *Pre-processing*: This stage includes data reading and initial cleansing, anonymization if needed, semantic annotation and formulation for processing; (ii) *Mining*: This stage includes processing tasks based on text mining tools and algorithms relying on a super-computing infrastructure, in order to produce service – oriented intermediate results.

The vision of ManyLaws is to produce and build the proper environment of semantically annotated Big Legal Open Data, one that is easily searchable and exploitable with proper visualization techniques. The ultimate objective is to provide the technical foundation and the tools for making legal information available to everybody, in a customizable, structured and easy to handle way.

The developed services will ensure real time provision towards citizens, businesses and administrations based on the most common needs of each user type. The current envisioned services provide the following: parallel search in many EU member-state legal frameworks using simple keywords (through parallel translation of search terms), assessment of the degree of transposition of an EU directive in a national legal framework, indicating relevant national legislation and monitoring the status of transpositions, analysis of references to the european legislation by national laws, comparative analysis of equivalent or relevant laws from different EU member states, comparative analysis of connected laws from the same member state, timeline analysis for all legal elements, visualising the progress and current status of a specific national or European legislation (after amendment/extensions) over time including preparatory acts and agreements, interrelation of laws and news or social media posts, including sentiment analysis, various geo-related visualisations (e.g. EU maps indicating different parameters), various text-related visualisations (e.g. wordle, sentiment graphs, interrelation maps, etc.) and other common visual aids (e.g. graphs, charts, tables, etc.), visualizations of correlations, dependencies and conflicts between different laws and decision support services (e.g. impact assessment) within legal procedures.

Although legal information is generally considered to be at the core of the open data movement and a major part of public sector information [1], there are implications in

considering the way text mining tools are used to produce results. Within this context operative services of Many Laws will facilitate not only the decision making processes but also enable access to legal information across the European Union. The legal and ethical considerations pertaining to the application and use of data mining tools and methods must not be overlooked. Wahlstrom et. al. [2] argue that data mining as a process is not in itself ethically problematic; instead ethical dilemmas arise when the data to be mined is of a personal nature. Individual consent (or lack thereof) to data being collected and used is a further ethical consideration [3]. In this context, [2] identify four ethical issues associated with data mining that warrant further investigation: privacy, data accuracy, database security, and stereotyping.

In this workshop, we will enquire into the need for a legal and ethical framework concerning legal text mining and, if so, the possible versions of this framework. Particularly we will determine the concerned data protection, copyright laws and find possible solutions for those concerns.

1.1. Relevance of the Workshop to the JURIX Conference

The Many Laws project aims to apply information processing and text mining tools and methods to big legal data in order to develop novel legal information processing services for citizens, businesses and public administrations within the European Union. The proposed workshop will focus on the factors that can contribute to the effective delivery of the new services, together with the legal and ethical implications associated with the application of advanced computing technologies to the acquisition, storage, and processing of legal information.

1.2. Questions to be Addressed During the Workshop

The main goals of this workshop are to identify the factors that enable the effective delivery of legal data-related services, and to discuss the key legal and ethical considerations associated with the application of information processing tools to legal text mining and database compilation. The following questions will be used as a starting point to guide group discussions:

- What are the factors against which existing projects concerned with the publication and access to legal information can be assessed?

- What are some of the key lessons from current legal text mining projects (eg. OpenLaws) that can be taken as starting points for Many Laws?

- Who are the key users of legal text mining services? What are their immediate needs and how can a project like Many Laws fulfil these?

- What are the central legal and ethical considerations when developing a legal text mining service based on AI and supercomputing?

- What are possible solutions for the legal and ethical challenges arising, and how could they be implemented?

2. Format of the Workshop

This half day workshop (3 hours) aims to bring together legal scholars, practitioners, and policymakers within a stimulating environment to discuss current developments, opportunities and challenges relating to the application of text mining, the compilation of databases, and the use of AI and supercomputing to facilitate the publication of and access to legal information in the European context. The workshop will be divided into two sessions, each comprising of three interrelated activities: 1) Introductory short presentations by the organisers 2) Group breakout sessions focusing on key issues, and 3) Interactive discussions to generate conclusions.

Ideas and insights generated through discussions during this workshop will be recorded by the organisers and shared with workshop participants and the general public at a future date. The structure of the workshop is outlined below in Table 1.

Table 1. Workshop Agenda

Activity	Time
Open Remarks and Introductions	10 minutes
Session 1: Legal Text Mining – Existing Tools and Infrastructure	
Presentation I: ManyLaws – Technical requirements and Proposed Infrastructure	10 minutes
Presentation II: Legal Text Mining – Existing Projects, Tools and Infrastructures	10 minutes
Breakout Session: The Value of Applying Information Processing Tools to Enable Access to Legal Information	30 minutes
Group Feedback and Discussion	30 minutes
Session 2: Legal, Privacy, and Ethical Implications of Legal Text Mining	
Presentation 3: Exploring the Legal, Ethical and Privacy Implications of Legal Text Mining	10 minutes
Breakout Session: Benefits of and Challenges to the Use of Text Mining within a Legal Context	40 minutes
Group Discussion and Final Evaluation	30 minutes
Formulation of Conclusions	10 minutes

References

[1] European Parliament, Revised DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information Directive 2003/98/EC sector information.

[2] K. Wahlstrom, J.F. Roddick, R. Sarre, V.Estivill-Castro, D. deVries, On the ethical and legal implications of data mining, Technical Report SIE-06-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia, 2006.

[3] L. Van Wel, L. Royakkers, Ethical issues in web data mining, Ethics and Information Technology 6, 2 (2004), 129-140.

Presenters' Biographies

Dr. Shefali Virkar is a Senior Researcher at the Department for E-Governance and Public Administration at the Danube University Krems (DUK), Austria, specialising in the theory and practice of electronic government. In particular, she studies the political, social, and economic implications of the new Information and Communications Technologies (ICT), and how these developments impact traditional forms of work and governance structures. She is has been involved on a variety of Austrian and European projects, and is a member of the EGOV-CeDEM-ePart conference organising committee. Shefali holds a D.Phil. (Ph.D) degree from the University of Oxford (UK), where her doctoral research focused on the impact that the perceptions and behaviour of political actors involved with strategic ICT projects in bureaucracies have on the ultimate outcome of such initiatives through an in-depth examination of a case study based in India. Prior to this, she was awarded a M.A. in Globalisation and Development from the University of Warwick (UK).

Mag. Anna-Sophie Novak is a Research Assistant at the Department for E-Governance and Public Administration at the Danube University Krems (DUK)**.** Anna-Sophie holds a master degree in law (University of Vienna) and is currently a PhD candidate at the University of Vienna.

Dr. Charalampos (Harris) Alexopoulos is a postdoc researcher at the Department of Information and Communications Systems Engineering, University of the Aegean. He is a Project Manager at the Information Systems Laboratory of the same department, working on European and National funded research and pilot application projects for governments and enterprises. Harris also serves as Programme and Organisation Committee Member, Track and Minitrack chair for Samos Summit, HICSS, MCIS. He is also a course manager of two summer schools and he is teaching e-government and business management at pre-graduate and postgraduate level. His research interests lie on the fields of Decision Support Systems, Open Data, e-government and interoperability. Harris is a computer science graduate from the University of Peloponnese with an MSc in Management Information Systems from the University of the Aegean. In 2015, Harris was ranked as one of the most prolific researchers in open data research worldwide by Hossain, Dwivedi and Rana (2015).

Dr. Yannis Charalabidis is Associate Professor in the Department of Information and Communication Systems Engineering of the University of Aegean. In parallel, he serves as Director of the Innovation and Entrepreneurship Unit of the University, designing and managing youth entrepreneurship activities, and Head of Information Systems Laboratory, coordinating policy making, research and pilot application projects for governments and enterprises worldwide. He has more than 20 years of experience in designing, implementing, managing and applying complex information systems as project manager, in Greece and Europe. He has been employed for 8 years as an executive director in SingularLogic Group, leading software development and company expansion in Greece, Eastern Europe, India and the US. He has published more than 200 papers in international journals and conferences, while actively participating in

international standardisation committees and scientific bodies. In 2016 he was nominated as the 8th most productive writer in the world, among 9500 scholars in the Electronic Government domain, according to the Washington University survey. He is Best Paper Award winner in the International IFIP e-Government Conference (2008,2012, 2016), winner of the first prize in OMG / Business Process Modelling contest (2009) and 2nd prize winner in the European eGovernment Awards (2009). As of August 2018, Yannis is among the 100 most influential people in Digital Governance worldwide, according to the apolitica.co list.

Mr. Michalis Avgerinos Loutsaris (male) is PhD candidate in the University of the Aegean. Michalis holds a Bachelor Degree and a Master of Science in "Technologies and Management of Information and Communication Systems" from University of the Aegean, Department of Information and Communication Systems Engineering. His research interests involve around Entrepreneurship, Artificial Intelligence and Text Mining. He is Lab Assistant and has some experience in teaching lab courses as assistant. He is member of the Information Systems Laboratory (ISLab) and is working in several lab projects.