



CREATE Working Paper 2014/11 (October 2014)

Report on a computer assisted copyright reform observatory

Authors

Ermo Täks
Tallinn Institute of
Technology
ermo.taks@ttu.ee

Addi Rull
Tallinn Institute of
Technology
addi.rull@ttu.ee

Anni Säär
Tallinn Institute of
Technology
anni.saar@humanrights.ee

Burkhard Schafer
University of Edinburgh
b.schafer@ed.ac.uk

This report contains research that formed in a much reduced form the basis of a contribution by the authors to "Logic and Legislation", edited by Michał Araszkiewicz and Krzysztof Pleszka, scheduled to appear in the Springer Legisprudence Library in 2014

CREATE Working Paper Series DOI: 10.5281/zenodo.12431.

This release was supported by the RCUK funded *Centre for Copyright and New Business Models in the Creative Economy (CREATE)*, AHRC Grant Number AH/K000179/1.

Abstract

In the past, computational representations of copyright law were almost exclusively used for DRM technology, and with that at the consumer side of the creative industry. The problems of this approach are well known. This project explores more "creative" approaches to computational copyright law - instead of focussing on consumers, it aims to utilise "self-applying" law to reduce costs both for the legislative process and also for the management of licenses and contracts by the rights holders and their legal representative. The paper proposes an approach to AI assisted law reform, that tries to align research in Artificial Intelligence and Law with the jurisprudential philosophy of Luc Wintgens. Taking a holistic, system-oriented view, we propose a visualisation based link analysis that allows lawmakers to identify those parts of the legal system where the smallest amount of change has the largest effect.

Keywords: Link analysis, Natural language processing, law reform, copyright law

1 INTRODUCTION

Adoption of Artificial Intelligence (AI) technologies in commercial legal practice remains very low. In a 2003 study of sixty-six law firms, located in Australia and Norway, it was found that expert systems development scored close to the lowest level on their scale (Apistola and Lodder 2005). Beyond technological limitations Oskamp and Laurits postulate broader reasons for the absence, and attribute a major factor to a poor understanding of what lawyers want. They call for a two way dialogue between legal practice and AI research, writing:

“until practicing lawyers see clear and immediate benefits of specific applications they will be reluctant to use them.” [11 sec 4]

The position of copyright law in this assessment follows with two important exceptions a similar pattern. Until very recently, it was one of the more under-served fields of legal AI research. We analysed 15 years of contributions to the proceedings of ICAIL and JURIX as the main conferences that serve the AI and Law community, and all 21 years of the Journal for Artificial Intelligence and Law as the main journal in that field. We also carried out a keyword search via Google Scholar, DBLD and CiteSeer to complement the survey.¹ Copyright emerged only recently as a topic of interest, and in particular did not play a role in the “classical” period of legal expert systems that model judicial reasoning.² In line with the general experience of limited take-up by legal practice, there was no evidence that AI technology had played a transformative role in copyright licensing or litigation, though both will have benefited like all fields of legal practice from “generic” computational tools for tasks such as information retrieval, case management or compliance assurance.

¹ Search string (Boolean) ““copyright law” AND “artificial intelligence” OR “knowledge engineering” OR “knowledge representation””

² by contrast, patent law makes a much earlier, and much more prominent appearance, undoubtedly driven by the prevalence and importance of patent databases

The first prominent contribution that we were able to find was the formal representation of an upper level ontology for copyright, which however constituted only one part of a larger investigation into formal models of Intellectual Property by Contissa and Laukyte (2008). Their paper reports mainly problems that were encountered in the process of developing a formal ontology for IP law. Generally, the emergence of formal ontologies as a building block of the semantic web has given a major boost to the research into formal representations of copyright law, and our own study follows this trajectory. Formal ontologies play a central role for the semantic web. The prevalence of license statements attached to digital objects in the digital economy yielded the first exception to the general rule that research in AI and Law rarely leaves the confines of academic conferences, proof-of-concept prototypes and feasibility studies. However, users of the technology are not the traditional targets of legal AI research – law firms, courts and public administrations – but Internet publishers of all hues. Creative Commons for instance has made licenses available in RDF (“Resource Description Framework”) format, which allows web publishers to embed license information in machine readable format in web pages, documents and mp3 files. Building on the notion of semantic web representations for digital resources and incorporating them into other knowledge management tasks this emerged as one of the most promising avenues for computational copyright law. An overview of these approaches, and also an indication of the uptake this research, if not by lawyers then by web publishers, can be found in Bourcier et al (2010).

As we have seen, after a long period of neglect computational representations of copyright law very recently garnered attention in the AI and Law community. The uptake by practice, in particular legal practice, remains however limited. There is one other approach to computational copyright though whose impact on practice and its commercial relevance can't be doubted, even though not all of the initial higher expectation have been achieved, and in some sectors of the creative digital economy, their use has been in decline [see e.g. Becker 2003; Ashtar 2010]. Digital Rights Management (DRM) “imitates” legal regulations by regulating access to digital resources. For Lessig, this new way to integrate legal compliance directly into software architecture became the archetypical example for an entire new form of Internet governance, regulation through software code. Despite this obvious connection between DRM and representation of legal concepts, interest in the AI and Law community has been muted. Only one paper, in the resources surveyed, attempts to leverage existing AI and law solutions for better DRM (Garcia et al 2007). A small number of follow up studies citing this initial study exist, but are published outside the core outlets for the AI and law community, such as (e.g. Prandoni 2009, Zhang 2012). The reason for this limited response lies in the different design philosophies in traditional AI and Law research on the one hand, DRM on the other. AI and Law research tries to make the logical deep-structure of legal knowledge explicit, resulting ideally in isomorphic correspondence between the formal model and the natural language legal reasoning (Bench Capon and Coenen 1992). DRM by contrast only mimics the consequences of legal norms, but does not give an explicit account of how these results are reached based on a fully explicit and transparent reasoning process. This has considerable advantages for the development of commercial applications: Since legal knowledge is not explicitly represented there is no knowledge acquisition bottleneck, a persistent problem in expert system design. There are also advantages for scalability and efficient use of computational resources: fully explicit legal reasoning can make considerable demands in terms of runtime on the computing environment – having a system perform an

explicit legal analysis, even if automated and supported by significant computing power, every time a digital photograph is opened or an e-book downloaded would mirror in the real world a situation where we ask for full legal advice when performing the most trivial of transactions. On the other hand, a recent decline in the use of DRM, and a focal point for widespread criticism of the technology is its frequent overreach (Favale 2011). It prevents transactions and operations the buyer of a digital object would be legally entitled to perform. This is the dark side of the same coin – absent an explicit and fully formulated representation of the legal environment, “dumb” DRM can catch only a small aspect of the legal meaning of “having a license” or “buying a file”. DRM’s relation to copyright is similar to that of a traditional lock to property law – a very rough approximation, but in the same way in which a physical lock will prevent entry also for people who are entitled to, so does DRM often prevent legitimate uses.

We face as a result a dilemma: DRM is “dumb” but has a track record of application in practice, even though one marred by constant criticism. Traditional legal AI is intelligent, and in principle capable of addressing the problem of “overreach” in DRM, but struggles to develop application of commercial strength.

There is a third computational approach that has the potential to change the nature of litigation in general and that of copyright litigation in particular. Copyright litigation can be data-intensive, for instance when investigation peer to peer networks which illegally share copyrighted material. AI supported e-discovery, especially in the form of advanced information retrieval methods, helps to manage the amount of information that has to be processed, and to reduce the costs in finding and curating the relevant data (Staudt 2003; Conrad 2010; Hogan et al 2010). The benefits of AI supported e-discovery, in particular those of predictive coding, have by now also been recognised by the judiciary (Barry 2011; Baron 2011) However, these approaches typically rely on statistical methods are even further away from the explicit modelling of legal concepts than DRM. The type of data that are typically the subject of e-discovery are not normally legal documents, so there is no need to optimise them for the identification of legal terms or connections between them.

In this study, we try to improve on existing approaches to computational support for copyright lawyers by combining ideas and experience gained across all three approaches. From traditional legal AI, we take the focus on legal texts and materials and the interest in the logical and syntactical deep structure of text. From e-discovery, we take the use of statistical methods as used in natural language processing. From DRM finally, we take the overall aim – lowering the costs of litigation and law compliance by automatizing some of the processes that underpin the functioning of the legal system. However, we also suggest a radical re-orientation of the efforts of AI in copyright law – away from the customer (or suspect, as the case may be) and towards legislators, law firms and legal in-house council as two sides of the same coin. Optimising legal regulation on the basis of sound empirical evidence – one of the core aims of CREATE – is inevitably disruptive and thus costly. We aim to demonstrate that a formal model of copyright legislation, developed using natural language processing methods, can optimise the process of law reform, and also assist law firms or inhouse legal departments that curate contracts and licenses to adjust more efficiently to the legal change.

This paper proposes a new approach for AI support for legislative drafting. Unlike many existing approaches, it takes a principled “sceptical” position towards legislation. Considerable increases in the quantity of legislation, and increase in the speed of legislative reform, create considerable compliance costs for businesses. Being intrinsically disruptive, they also create problems for law firms and legal counsel, for instance in the form of training costs. Our proposed system, which we discuss using the example of copyright law reform, takes an outward facing and an inward facing perspective. In the inward facing perspective, it addresses the legislator. The aim is to assist him in identifying “minimally disruptive” ways to change legislation. This can mean either to achieve a considerable degree of substantive legal reform by changing only a small number of “central” laws – the aim of a major reform project – or restricting the change as much as possible by identifying suitable candidates for reform that are only minimally interlinked with the rest of the legal system. In this we take our inspiration from link analysis, in particular the analysis of criminal networks and the way in which the results inform police practice when criminal networks. In the “outward facing” mode of the system, it assists lawyers at predicting the impact a proposed law reform is likely to have on their business and clients, e.g. by estimating how many contracts or licenses need revising in the light of the legal change, or by identifying and triaging training needs. The aim is in both cases to reduce the costs of litigation by reducing the costs that are created through disruptive legislative activity.

Assisting legislative drafting has over the past two decades become a major research interest within the AI and law field (see e.g. Voermans and Verharen 1995; Winkels and Den Haan 1995; Breuker et al 2000; Hokstra et al 2003; Hafner and Lauritsen 2007; Palmirani 2011; Voermans et al 2012) Over the same period of time, legisprudence, the jurisprudential analysis of legislation, also rose to prominence. While the term “legisprudence” is slightly older, having been coined by Julius Cohen in 1950 (Cohen 1950), it was revived only more recently, with the publication in 2003 of the Proceedings of the Fourth Benelux-Scandinavian Symposium on Legal Theory, under the title “Legisprudence: A New Theoretical Approach to Legislation”. In 2007, a new journal, *Legisprudence*, was founded by Luc J. Wintgens and published by Hart. In 2008, Boston University’s law school held a symposium on legislation, the proceedings of which were published in the *Boston University Law Review* in 2009. The term “legisprudence” is by then used as a matter of course, most notably in the *Editors’ Note on “Legisprudence”*, 89 B.U. L. Rev. 331 (2009), which notes that “Legisprudence has a short history, but a long tradition.”

AI approaches to legislative drafting and legisprudence share ostensibly the same goal – theoretically informed improvements of the legislative process. Despite this convergence of interests, there has been so far little systematic reception of legisprudence research in AI and law. We argue that this is not by chance, but that there is a real if non-obvious tension between the approaches favoured by AI researchers, and the positions advocated by some of the most influential legisprudence researchers, in particular the work of Luc Wintgens and Jeremy Waldron. In response, we propose a somewhat different way in which artificial intelligence can support not just the legislative process, but also law firms as “consumers” of legislation. By taking the notion of a legal *system* as our starting point, we conceive law similar to a social network, with different parts “communicating” with others in varying degree. By matching and reengineering these “channels of communication”, we can transpose other concepts from social network analysis to assist in a way to reform law that is potentially better aligned to then theoretical commitments of legisprudence.

2 THE INHERENT TENSION BETWEEN AI AND LEGISPRUDENCE.

In this section, we explain in more detail where we see the inherent tension between AI approaches to legislation and legisprudence. The information technology revolution has often been called a second industrial revolution (Forrester 1985). As with the industrial revolution of the 19th century, this mainly means the promise to produce goods or services faster, therefore cheaper, and therefore more of it to satisfy the increased market. “More, faster and cheaper” are also promises of the ICT revolution (see e.g. Becchetti et al 2003; Swierczek et al 2005) For AI and law, this is particularly visible when systems are designed to increase access to justice: if courts could resolve disputes faster and cheaper, people currently excluded from justice could get their day in court (and if it is a virtual one) which as a result would also deliver more decisions (see e.g. Berman and Hafner 1989; Staudt 2008). But are “laws” the type of entity where “more” ceteris paribus equals “better”? As we will see in more detail below, we have reasons to distrust this idea. One need not share Thoreau’s anarchism to see some truth in his dictum that “the government is best which governs least”. Legislation, as exercise of sovereign power, is inevitably also always an exercise in violence. Justified violence in many cases, but not necessarily something one should cherish as an end in itself. The deeper philosophical issues that increased “legislative productivity” can raise were highlighted in Lon Fuller’s influential book, *The Morality of Law* (Fuller 1964). There, Fuller develops the parable of the inept lawmaker King Rex, who despite his best intentions fails to actually make law. In his 8th attempt, he increases the efficiency of his lawmaking to new heights, changing (and maybe even in a sense “improving”) laws on a daily basis. But even if the laws were objectively improvements and of good quality, the sheer speed of reform made their actual implementation impossible, and thus they missed their function to “subject human conduct to the governance of rules”, as the citizens could not any longer determine what the operative rules were.³ While this seems a farfetched example, typical for philosophical analysis but far away from the reality of law, at least anecdotal experience indicates that contemporary lawmakers can encounter similar problems. During the BSE crisis in the UK, the government issued almost daily “statutory instruments” that clarified and concretised the relevant laws as new medical evidence and data became available. When academics tried to write up the relevant developments, at least one judge approached them asking for further support, as he had decided cases on the basis of rules that had by that time already been revoked and faced the general problem of not receiving the updated laws in a fast enough.⁴ A slow and cumbersome legislative drafting process by contrast helps to generate the necessary stability that the application of law requires. It also helps to avoid undue public pressure after particularly high profile tragedies or scandals. Emotionally charged atmospheres, e.g. after high profile crime, are not necessarily an environment where rational deliberation about law can take place, “hard cases” often result in bad law (Brazier 1997). If the legislative process results through complex procedure in inevitable delays, a distance to the events is created that can result in a more sober analysis of appropriate legal responses.

³ Fuller considered the relative stability of law an aspect of its „inner morality“. Herbert Hart, famously, disagreed, calling it a mere functional requirement of efficacy (Hart 1957). For our purpose, the precise classification is irrelevant, what matters is that rapid change in the law is a problem for the legal order

⁴ Scott Wortely, in personal communication. The background to the problem can be found in Aitken 1997, which also illustrates the speed with which rules were created and repealed at that time

So far we discussed just the question of speed, tentatively concluding that computer support to increase the *speed* of law making is at the very best an ambivalent achievement. The speed with which things can be produced is however only one indicator for productivity. Another diagnostic criterion is a reduction in costs – producing the same amount of a product, but at a lower price. Translated into a legislative setting, this could be achieved e.g. by reducing the number of people involved in the legislative drafting process, or to supply at lower costs the expertise that is needed to inform the content of the regulation. Another option yet is to increase the range of products that are produced, while keeping the costs and the rate of production constant. To address the question whether a) these ways to increase productivity are unequivocally beneficial if the products are “laws, we need reflect further on what we mean with a “good” process of law making. As this is the main concern of any theory of legisprudence, it is there that we turn for help, looking in particular at the legisprudential theory of Luc Wintgens as described in his 2006 paper *Legisprudence as a New Theory of Legislation*. In this paper, For Wintgens, any theory of legisprudence as a rational theory of legislation has to start with “a reflection on the organisation of political space since Modernity”, and with that inevitably with a reflection on freedom (Wintgens 2006 p. 2). Wintgen’s pursues a twofold strategy. The first part of his paper is a critical account of traditional jurisprudential theories. His main criticism is directed at Hobbes and Rousseau, but they are only used as stand-in for all contemporaneous attempts to develop a theory of law and the state that is “scientific” in the way they inherit this notion in turn from Descartes. For Descartes, true jurisprudential statements, just like any other true statement of practical philosophy, must secure their certainty by being derived from first principles that have the same clarity and distinctiveness as the famous *cogito ergo sum*. From such a certain and foundation, reality can be “build up” in a rational way. As a corollary, everything that is not amenable to logical or empirical proof is not rational and hence also not scientifically sound, true knowledge.

Here we encounter a problem that also has implications for the project of AI assisted law making. Values and preferences of individual actors are not capable of logical or empirical proof; In the thus reduced conception of knowledge sketched out above, that means that they are as a result not rational, and hence not subject to true knowledge. Hobbes Rousseau and the other contract theorists resolve this issue by basing their theory of the law on the concept of the social contract.

Just like the *cogito* in theoretical epistemology forces its truth on us, so does , reflecting about ourselves and making use of our rational capacities lead to the inevitable conclusion that entering into the social contract is preferable to staying in the state of nature. This is particularly clear in Rousseau: it is reason itself unfolding that leads to the “true principles of public law”. The truth of the premise, i.e., the social contract, logically leads to the truth of propositions based on it, i.e., laws. This, as Wintgens argues, leads in turn to legalism, the pattern of legal thinking that was dominant from the seventeenth through to the middle of the twentieth century. Normative behaviour is then reduced to rule-following (Shklar 1964, 1), and with that also a conception that it does not really matter where these rules come from (Bankowski 1993). Law is “just there” (Wintgens 2006).

Once legalism is established as the necessary consequence of modern theories of state and law, Wintgens lists four further characteristics of the legal system thus conceived: First, if the construction of laws results in true normative propositions (because they can be derived from the social contract) , Since truth, in the Cartesian tradition, such propositions must be

timeless. Secondly, and most importantly for our issue, since laws are true there can be no scientific discussion about their content. To quote Wintgens:

“This entails that the disputable nature of values, goals and ends is concealed. Any rule is true which means that the value, goal or end is morally correct. On this view, laws are considered instruments for their realisation without any need to be chosen. This characteristic of legalism can be called concealed instrumentalism.”

The third factor states that once the social contract is concluded, any normative proposition of the sovereign ipso facto trumps any other normative statement. Since the law of the state describes what is right or wrong, “merely” subjective moral intuitions of the subject become irrelevant.

Fourthly, “legal knowledge” or study of the law is the knowledge of the true legal propositions. Consequently, the legal system is a closed set of logically connected propositions. A science of law is possible provided it restricts itself to the analysis of law as always already given by the sovereign.

Why does this matter for a research project in AI and law, or more precisely, AI and law making? “Strong Legalism” of the type Wintgens describes as the dominant paradigm of (secular) legal science provides an obvious foundation for legal AI, understood as formal models of legal reasoning. But at the same time, it renders any attempt of *legal* AI in the service of law making impossible, or so we argue. If law is nothing else but a closed set of logically connected propositions, possibly closed under deduction, then legal expert systems are as easily designed as early proponents of this idea thought they would be – an inference engine and a list of true propositions in the knowledge base is all it takes. But by the same token, and for the same reason, a legal expert system that assists law making is a contradiction in term. Legalism excludes any form of theorizing about legislation. Quoting again Wintgens (2006):

“Legislation is a matter of politics, and politics is a matter of choice. Choices are disputable, so a theory that would take them to be the object of knowledge is condemned to failure from the very beginning.”

Knowledge of the rules is both necessary and sufficient to know what we ought to do, they fully describe all rights and duties. The first stage in building a legal expert system is knowledge acquisition (Boose 1989). In legal AI, as in AI in general, this will typically involve also an evaluative stance – what is the best practice in a given domain that the system can or should model (O’Leary 1998)? But we have just proven that there can’t be such a thing as “knowledge” of a good legislative process, let alone one that can be modeled through rules! For if there were such a thing, the legislator as a sovereign actor within political space would be bound by rules, and if he were, he would not be a sovereign. Judges by contrast can be bound by rules, which quite naturally results in the reduction of jurisprudence to the theory of the application of rules by judges.

Since law-making is not a matter of legal theory, there can’t be a theoretically informed computational modeling of law making either. For sure, there can be other ways to assist a lawmaker through IT support, but the best we could hope for are generic approaches that would equally benefit the smooth running of any organization, issues such as document management and information retrieval for instance, but none of these would be a genuinely

legal AI approach. If we look at the history of AI and law, we can indeed see this idea reflected, AI and law, for much of its history, behaves as if strong legalism were true. The earliest rule-based expert systems such as TAXMAN (McCarthy 1980) or Sergot's influential formalization of the British Nationality Act (Sergot 1989) most clearly mirrored a conception of law as sets of proposition closed under deduction, and while subsequent approaches enriched and refined this model substantially, the main focus remained an attempt to find those rules that mirror most correctly the application of the law by a judge at a give moment in time. By contrast, interest in the legislative process is a more recent interest, and where it does come to the fore, it is indeed often through systems that are largely "content neutral" in the sense that they could support any complex administrative task, and also are much less likely to align themselves to specific debates in jurisprudence the way in which this is not uncommon in gel expert system design.

Legal AI that tries to improve the legislative process therefore faces several challenges: If understood merely as generic IT system design applied to the legislative process, it may be capable to increase the speed or reduce the costs of lawmaking. But even from a purely utilitarian perspective, simply improving speed or cost of the legislative process might be a mixed blessing. The inevitable time delays allow legislators necessary distance to highly charged and emotional problems, the costs give them incentives to use legislation only as a last resort. Removing or reducing either obstacle may result in even more juridification (Teubner 1987) of the social world, even more rules and regulations that create burdens on the norm recipient with often little benefits for society (Blichner, and Molander. 2008). If our aim is more ambitious though, and we hope to develop AI approaches that are build on specifically *legal* knowledge and assist in better or more varied outputs of the "product law", then we face the problems discussed above: if legalism is true, then there can't possible be the type of knowledge in rule form that could underpin such an endeavor, and no legal theory, qua legal theory, that could provide the theoretical underpinning. If legalism is false though, AI and law in general face problems.

Wintgens however offers us more than a mere critique of legalism in the tradition of Rousseau or Hobbes. He proposes instead what he calls "weak legalism", and with that an option to "bring in" a theory of law making under the umbrella of legal theory that stays true to modernity's understanding of law and the legal process.

In this alternative model of "weak legalism" freedom is asserted as a general principle of the legal order, an a priori goal that al laws have ultimately to serve. This creates a justificatory pressure on the social contract and the laws that it generates. Citizens, upon entering the contact, do not give a general and irrevocable authority to legislate on their behalf to the sovereign. Rather, in every single instance, and for each and every external limitation of freedom through law, the trade off between values such as life and safety has to be balanced against he loss of freedom. Where in the Hobbesian model, the formation of the social contract results in a general and *a priori* trade-off of freedom, citizens in the model of weak legalism retain in principle their moral capacity to act on their conceptions of freedom. Once freedom is established as the governing principle of legal order, external limitation of freedom must be justified in the individual case. For Wintgens the realm of legisprudence is precisely to be a theory about the permissible justifications and their evaluation. The duty of justification is what legisprudence is about. As a rational theory of legislation, it provides the general – and hence rule based – principles that allow us to create legal orders that maximize freedom through justifiable trade-offs.

Such a principled framework that enables the justification of external limitations, and with that the process of their legitimation, in turn would provide us with the type of knowledge we need to construct Artificial Intelligence tools in its aid. Wintgens suggests four principles in particular. We will try to show how they can be understood also as a form of “requirement engineering” for legal AI in the legislative process, that is as part of the process of “soliciting, structuring and formulating software requirements” and so “a systematic way of producing system models” (Sommerville and Kotonya 1998 p.139). We will therefore briefly discuss all four principles as defined by Wintgens and indicate some of the consequences we can draw from them for the design of our software,

The four principles that an AI enabled legisprudential drafting aid should model are:

- the principle of alternativity (PA)
- the principle of normative density (PND)
- the principle of temporality (PT),
- and the principle of coherence (PC).

The Principle of Alternativity (PA)

The Principle of Alternativity is the most abstract of the four principles and to a degree encapsulates and summarized the previous discussion. It establishes as default that as long social interactions work without external guidance through laws, there is no legitimate role for legislation. Legislation is the answer to failing social interactions only.

In the theory of Internet governance, Larry Lessig popularized the regulatory square of legislation, market regulation, social norms and architecture. For Lessig, these are in principle equally valid ways of guiding behavior, though different situations may make on or the other the more efficient way of intervention. Since legisprudence however concerns itself exclusively with the regulation through for laws, its underlying assumption is that social practices are self-regulating. Citizens create meaning through their interactions, and in doing so refer to rules that are embedded in social practice. Unlike Lessig, but following Riceur, this creates in Wintgen’s model an asymmetry between formal laws as method of regulation by the sovereign and other forms of regulation. Citing approvingly Hunyadi (1995), his model too assumes that the existence of these rules becomes visible in the case of conflict (Hunyadi 1995). In conclusion therefore, PA claims that the sovereign can only intervene if he can show that

a) there are societal conflicts

and

b) other modes of social interaction fail to resolve them

Then and only then, external limitation is preferable to an internal limitation of freedom.

From the perspective of a *legal* AI support, PA poses a number of challenges. It requires a comparative evaluation of legal and non-legal modes of regulation, and therefore points with necessity beyond legal knowledge and its formal modeling. There has been recently considerable interest in “evidence based policy formulation”, and the type of questions that

this sort of exercise raises would also be the target of a computational model of PA. As a knowledge intensive task, IT based support seems prima facie possible, but the knowledge that would be modeled is not legal knowledge, but based on economical, psychological and organizational data. Examples of this type of knowledge in the field of copyright reform, the main application of this paper, can be found for instance in Towse (2011) or Kretschmer and Towse (2012).

Slightly more accessible to a purely legal analysis is condition a). We could for instance take the number of cases litigated under a law as a proxy for the fact that there is a social conflict to be resolved. A simple link analysis system that connects statutes to cases could then give us an indication of a specific legal provision “earns its keep”. There are admittedly problems with this assumption. The most obvious imitation is that it assumes that there is already a law in place whose usage can be measured. It fails in those situations where we contemplate enacting an entirely new piece of legislation. A law might be so efficient, or its deterrent so high, that no further litigation arises once it is enacted. Alternatively, a law that does not impute a sanction, and is formulated in a particularly precise way, is unlikely to create litigation. An example could be a rule that limits tort claims to 5 years after an accident has occurred. Apart from possible constitutional challenges, it is unlikely that a norm like this will create a significant amount of case law. Conversely, a large amount of litigation on a specific legal provision can well be a sign that it is badly drafted, not a sign that there is a considerable underlying social problem in need of regulation. We will come back to this issue when we discuss the principle of temporality, where we will argue that PA and PT together allow to a degree a computational assessment.

The Principle of Normative Density (PND)

The PN states that rules which impose sanctions need a special justification, and the more severe the sanction, the greater the pressure on justification and legitimation. The intuitive reason for this rule is the “double impact” of sanctions on the concept of freedom. A norm for instance that sanctions a certain behavior X with a prison sentence, first reduces freedom by limiting the right of the citizens to choose behavior X, and then in the case of rule violation reduces his freedom even further, by restricting his ability to act on his preferences in a significant way. In practical terms, this means that everything else being equal, and PA deemed to be fulfilled, a law that does not impose a sanction (but creates e.g. an agency charged with assisting subjects in following a norm, or creates a new power that helps citizens resolve the conflict amongst themselves) is preferable over one that imposes a sanction on the citizen.

Wintgens argues that the range of possible legal consequences come with a variable degree of normative density, with sanctions marking the maximum of impact on citizens. Norms that require mere information duties and thus enable informed consent by contrast could be considered as a minimum density rule. We agree that this is an important first step to quantify the burden of proof on the legislator when legitimizing a specific proposed norm. However, we also argue that it should be properly understood in the context of law as a complex system. “Density” then means not just the legal impact of one individual and isolated rule but should best be understood as the totality of possible legal sanctions that attach to a certain real life activity. This allows in principle for a situation where the density of regulation, understood now as the quantitative degree of regulation of a sphere of life, is extraordinary high, even if none of the individual rules carries a sanction. In some cases, this regulatory

density can be so high that a trade off against a smaller number of sanction-carrying norms can on balance increase freedom. (as an example, one could think of a highly burdensome regime of reporting even the most minor deviations from numerous “best practice standards” in a hospital setting as opposed to a norm that does indeed impose prison for medical malpractice, but only in cases where serious harm was actually caused). We think that this understanding of the rule of density is in line with Wintgen’s own emphasis on the systematic character of law, discussed below, even though it is not expressed in this way when PN is introduced.

The Principle of Temporality (PT)

PT emphasizes the temporal dimension of laws. As argued above, one problem with strong legalism was the notion that laws of man and laws of nature share an atemporal quality. Once our theory of gravity is justified by overwhelming empirical evidence, we only revisit it under exceptional circumstances. If we extend this idea to the realm of human laws, we create a strong presumption that existing laws, at least approximations of the “best” possible legal solution, remain valid until proven otherwise. PT takes the opposite approach. Laws are creation of human beings in reaction to perceived conflicts. What might constitute a reasonable trade off between conflict settlement and freedom in a specific context, at a specific point in time may well be an unnecessary intrusion into our ability to live our lives according to our own conceptions of a good life. In the field of technology regulation for instance, it can be sensible to restrict or prohibit the use of a new and untested technology initially, or to impose special liability or duty-of-care regimes. A classical example are “red flag” laws, named after the Locomotive Act that mandated that trains with more than 2 vehicles should have a man with a red flag walking at least 55 m ahead of each vehicle, to enforce a speed limit and warn horse riders.⁵ While a reasonable precaution while the technology was new, it soon became obsolete as people developed appropriate patterns of behavior when encountering a locomotive. PT stresses the historical character of rule and their justifications. In legislative practice, it points towards the use of “sunset clauses” as a drafting default, which require the legislator to revisit periodically the justification and efficiency of a law (see e.g. Davis 1981; Finn 2009; more critically Kysar 2011). From an AI terms, it indicates the need to revise and update the system regularly, incorporating mechanisms such as non-monotonic logic to model the necessary element of belief revision. Wintgens notes that “This process of justification should include the consciousness that external limitations must be kept in track with changing circumstances. Obsolete legislation or external limitations that are eroded by desuetude are no longer legitimated. They are to be withdrawn, changed or qualified in view of the PA and the PN”. While we agree in principle, it should also be noted that change, even justified change, comes with costs attached to it. We introduced above the example of King Rex, and his failure to legislate properly when his law reform agenda became too fast to be implemented. We will come back to this point when outlining the business model for CoReO, and note here only that PT as a regulatory aspiration faces costs that may at least be reducible through the use of technology.

The Principle of Coherence (PC): The Level Theory of Coherence

5

https://www.direct.gov.uk/prod_consum_dg/groups/dg_digitalassets/@dg/@en/@motor/documents/digitalasset/dg_180212.pdf

The final principle Wintgens proposes is the notion that justification of legislation needs to look at the legal system as a whole. According to him, a legal system is “a complex and dynamic set of intertwined propositions” so that any change in one part of the system may affect it as a whole. The exponential growth of the number of laws threatens this systemic character, as the degree of complexity, and with that the number of complicated inter-systemic interactions, becomes difficult to control.

3 TURNING LEGISPRUDENCE INTO AI SPECIFICATIONS.

From the above discussion, we can now consider how an AI system that assists the type of jurisprudential analysis of law making that Wintgens proposes should look like. Here and in the example in the technical part below, we use copyright reform for illustration purposes. Copyright law is a suitable test case for a variety of reasons.

- it is an area of persistent legislative activity over an extended period of time. From the statute of Anne in 1710 to the present date, every generation seems to have amended, modified or extended their copyright scope. This is a tendency that is largely independent of jurisdiction, and even in common law countries, much of the reform dynamics was driven by legislative intervention and not just by the courts. The result has been a complex legislative framework dominated by general rules and a plethora of more and more fine tuned exceptions.

- despite this, it is a field of law where there is strong evidence that underlying social conflict remains unresolved (the “broken copyright” meme, see e.g. Tehranian 2007; Kretschmer 2008; Samuelson 2012)

- law reform therefore remains on the agenda for the foreseeable future, as evidenced by the Hargreaves report in the UK (Hargreaves 2011) or the discussion on the “third batch” of copyright reform in Germany (Beger 2010; see also Party 2012 for the US).

- it is an area where alternatives to legislation are discussed prominently, for instance regulation through markets or, famously, through computational architecture such as DRM systems (Lessig 1999).

- regularly, pressure for law reform originates outside national jurisdictions, for instance through the Berne Convention for the Protection of Literary and Artistic Works or within the EU the Directive 2001/29/EC (“Copyright Directive”) from 2001. Monitoring foreign and international law can create particular burdens on law firms and increase costs of litigation.

We can now look at a specific legislative reform proposal, such as recent proposals in the UK to follow the lead of the US and establish an exemption in copyright law if an otherwise protected work is used for the purposes of parody (see e.g. Deazely 2010; Mendis and Kretschmer 2013). In an ideal world, an AI system for legislative support that embodies the above discussed jurisprudential principles would then help to evaluate the proposal against the five parameters, and on this basis determine which possible route would maximize freedom while minimizing social conflict. We can immediately see why this would be an overly ambitious task that goes well beyond the capacities of any current or near-future AI system. Would creating a new exception even require an evaluation against the Principle of Alternativity, and would it increase or decrease normative density? Looking only at the syntactic level, on the one hand a new regulation is created, the sheer “amount” of legislation

increases. This might trigger our evaluation process. But then, the new norm is an exception to an already existing rule, merely counting words in statutes could therefore be misleading, and we should possibly treat it more like a revocation of a law than a new enactment. Presumably, though Wintgens does not say so explicitly, mere revocation of laws do not require the same type of justification that enacting a new law does, even if in the legal system in question, amending laws or revoking them formally involves enacting a separate piece of legislation. The same ambiguity can be found if we drill down a bit further and look at the semantic content of the norm. In one reading, a possible sanction for an action is removed through the suggested reform. While in the past, I would have faced civil litigation and damages, or even criminal prosecution for copyright theft, had I used someone else's creative work for parody purposes without the creators permission, I'm now free to do so. From the perspective of both PA and PND, this seems prima facie to be a net gain in freedom. However, analyzing the problem like this hides some important political and philosophical choices. It treats copyright as a mere regulatory system that tries to enhance market efficiency. Another way to frame the problem could be from the perspective of the creators. If we think of copyright as a natural law that precedes the social contract, then the new exemption not only interferes with a right, it potentially also creates a regime of sanctions. As an artist, if I now act against the person who uses of my work for a parody, by withholding e.g. money owned to him or by pulping his offending works, I can in turn be subject to both civil and criminal litigation and sanctions. In this reading, PA and PND are highly relevant. Which of these two conceptions of copyright is "the better one" or prevails is a substantive question of policies discourse, not something that can be mechanically deduced from first principles through a process of computation.

We conclude from this discussion that while we cannot hope to give a computational reading of the full semantic content of PA, PND, PC and PT, we can nonetheless assist the legislative process by identifying suitable proxies for each of them. Ideally, they should be amenable to a quantitative treatment and be based on syntactic categories, rather than involving qualitative reasoning about word meaning. We already indicated one very crude way how PA could be translated into such a computational format. On the most basic level, we could count individual norms, and rate the justificatory burden relative to the degree of additional (or diminished) regulatory instruments that they create. A more refined version could distinguish between primary rules and exceptions to rules, even more refined, but still a question of syntax and a limited representation of the operative vocabulary could take into account of the rules have punitive sanctions, or are rather creating powers or privileges. However, the added value of such a computation friendly evaluation would be slim for all but the most complex pieces of law reform, while telling the legislators mainly things they know already.⁶ However, if we approach the problem mainly from the perspective of PC, another option becomes available for us. If our starting point is law as a complex system that at the point of legislation is "always already there" and of such a complexity that its interaction with a new piece of proposed legislation can't be any longer immediately be seen, then modeling these possible interactions tells us something about the impact the new law will have on PA, PN and PC. It tells us for instance if what is intended as a small, technical change to one

⁶ exceptions to this could be particularly complex and complicated pieces of legislation such as the Patient Protection and Affordable Care Act in the US, which had not been read by most of the politicians that voted for or against it. Here, even a syntactic parsing exercise that tells the decision maker how many new duties, privileges, exceptions etc are created might result in better informed judgments.

specific piece of law only could “percolate” through the entire system if this law is interconnected with other regulations in specific ways. This can then allow us to determine where a change in the law causes the greatest effect. Sometimes, we may want to cause a substantial reform of the entire legal system, ideally by enacting as few new laws as possible (because this reduces the costs for retraining lawyers, and also reduces the cost of the legislative process). Sometimes, we only want to affect minimal change, and are concerned that interfering with the system could have unforeseen consequences. In our example for instance, after the empirical data as been collected, we might conclude that the problems experienced by users of copyrighted material are but a side effect of a more general malaise of a country’s property regime. In this case, introducing one single law that creates a “social acceptability” exemption that governs all forms of property, from real estate to movables to intangible property, would achieve the maximum desired change with the minimum of legislative effort. Conversely, we might find that this type of dispute is typical for written work only, and therefore does not even require a general new exception for all types of copyrighted work. In this case, the least disruptive reform has to ensure that no other parts of the legal system are accidentally affected, for instance by introducing a new definition for a term that is also used elsewhere.

This type of cognitive operation is similar to that carried out by a police officer who contemplates how to deal with a criminal network. In some cases, it might be best to “take out” a highly connected member of the organization, in the hope that it will disintegrate quickly as a result. In other situations, arresting people at the periphery may be preferable, to prevent the main target from being alerted to the fact that s/he is subject to an investigation while minimizing social harm, or to prevent succession conflicts and inter-gang wars breaking out. Computer assisted link analysis and visualization tools have been shown their capacity to play an assistive role in this task (see e.g. Xu and Chen 2005; Schwartz and Ruselle 2009; Hutchins and Benham-Hutchins 2010).

The idea to represent the legal system as an interconnected graph like structure is by no means new, though the application suggested here, utilizing them for more strategic law reform, may well be. Typically, they base however their approach on explicit references and citations (see e.g. Bommarito and Katz 2010; Boulet, Mazzega, and Bourcier 2011; Kim 2013). For our purposes, this means that while they can and should play an important role for the type of system we envisage, they only give a limited picture. Firstly, they only map those connections that the legislator was aware of. But as our discussion of Wintgens above showed, the complexity of the legal system is partly due to its organic growth that eluded strategic planning. The legal system, as Wintgens argued is not merely complex, it is complicated. These complications are the result of unintentional, unplanned interaction between the constituent parts of the legal system, which gives particular prominence to PC as an aspect of a rational theory of jurisprudence. Secondly, what we are interested in is not (just) the explicit correlation between statutes, which follow the internal logic of the law and its administrative and historical subdivisions. Rather, following the analysis of PA and PND above, our concern is the regulatory density of “spheres of life” (“Lebenssachverhalt”). An artists faces in her creative work external, legal constraints that cross and combine a multitude of heterogeneous legal issues, from copyright to employment law to contract law to criminal law. The law divides and separates artificially issues that phenomenologically, from the perspective of the norm recipient, are but one set of external factors that limit their freedom. It is unlikely, and from the internal organizational logic of the law indeed

undesirable, to connect explicitly all the relevant laws through explicit references and citations with each other. If however we want to determine if there are alternative, if more remote, ways available to regulate a certain real life problem, we cannot limit ourselves to the “obvious” explicit connections between laws. Rather, we have to discover existing but implicit pathways between regulations that can impact on the same set of factual circumstances. Our approach maps therefore graphically those connections between laws that are not (just) the result of explicit citations, but are the result of the semantic features of the natural language expressions that the legislator chose to frame the law.

4 TECHNICAL IMPLEMENTATION

As discussed above, legal systems evolve incrementally over time. Even in post – revolutionary legal orders, law reform barely starts on a blank slate. Rather, we are always already confronted with a complex set of interacting norms that are for most people opaque. This insight from jurisprudence motivates our approach of “reverse engineering”. Rather than building a legal system from scratch, the way strong legalism uses the social contract model, we take the existence of complex legal systems that are at least partially impenetrable as given. “Reverse engineering is the process of developing a set of specifications for a complex hardware system by an orderly examination of specimens of that system. Following Chikofsky and Cross (1990), it is a process of analysing a system or complex object to:

- *Identify the system’s components and their interrelationships and*
- *Create representations of the system in another form or at higher level of abstraction.*

Restructuring is the term for creating a new system with qualitatively new characteristics, using the same building blocks but preserving the subject system’s external behavior (functionality and semantics). It can be applied upon legislation to achieve different purposes, to name a few of them: cope with complexity, generate alternative views, and synthesize higher abstractions.

From this point legislation is consisted from individual small particles, norms, which are grouped based on similarity and combined together in the form of legal texts. These structural units do carry certain semantic meanings, different norms within paragraphs for example are usually grouped together in order to regulate one specific legal response to a perceived societal need. A norm typically contains two parts, the description of the state of affairs that is regulated, and the legal consequence that is triggered if that state is met. From the perspective of the legal system, these consequences are a main organizational feature – we group e.g. all those norms that have a prison sentence in the consequent (the “then” part) of a rule together under the label “criminal law”. As indicated above though, for our purposes the clause in the antecedent of the norm (the “if” part) is at least as important to identify structural connections between norms, so that in our view, a rule that decrees a punishment for an artist who libels through his work a person, and a norm that enables an artist to recover monetary payment for his work, are connected even if they come from entirely different legal fields. The clause is the smallest linguistic structural unit found able to carry meaning of one norm. The clause is the natural container able to represent full meaning of the norm by itself and therefore is chosen to be the basic unit of legislation structure analysis.

By definition the clause is a group of words containing a subject and predicate and functioning as a member of a complex or compound sentence. (Clause) The subject is a noun and the predicate is a verb (Definition: Subject, Predicate). Simplifying a bit and in order to estimate the most meaningful, most information loaded elements of the norm, so call “data carriers”, we can focus only on these two types of words. Our experiments have shown that picking out only these two types of words it is possible to capture 60...80% of the norm content embedded into clauses.

Such an approach allows us to extract and visualize the norm content. Linkage between words can be presented as a graph; relative importance of the link can be illustrated with help of connection “weights”, which in the current case reflect the frequency of use these word within the clause. Computationally this means the formation of two-way tables, consisting of nouns (in rows) and verbs (in columns) with the frequency of found word pairs the evaluative criterion. Verb and noun concordance two-way table representation can be used to capture a concentrated content of the single norm but also for the set of closely related norms or groups of norms, represented in text as a paragraphs, chapters, legal acts or even entire legislation.

As an example we take a norm from the Estonian Constitution, § 20:“*Everyone has the right to liberty and security of person.*” With a use of suitable text analysis software it is possible to automatically estimate the borders of the clause, the type of words (nouns, verbs, numerals, conjunctions) and count the frequency of noun and verb pairs presence as shown in Table 1.

Table 1. Two way table representing the content of the norm (Täks, Kuusik, & Nyman-Metcalf, 2013)

Noun/verb	has
everyone	1
right	1
liberty	1
security	1
person	1

Further scaling of this method toward larger sets of captured norm contents will bring us to more sizeable cross tables and graphs, a small examples of which is shown in the next example (norms are extracted from Estonian “Parliament Election Act”), in table Table 2 .

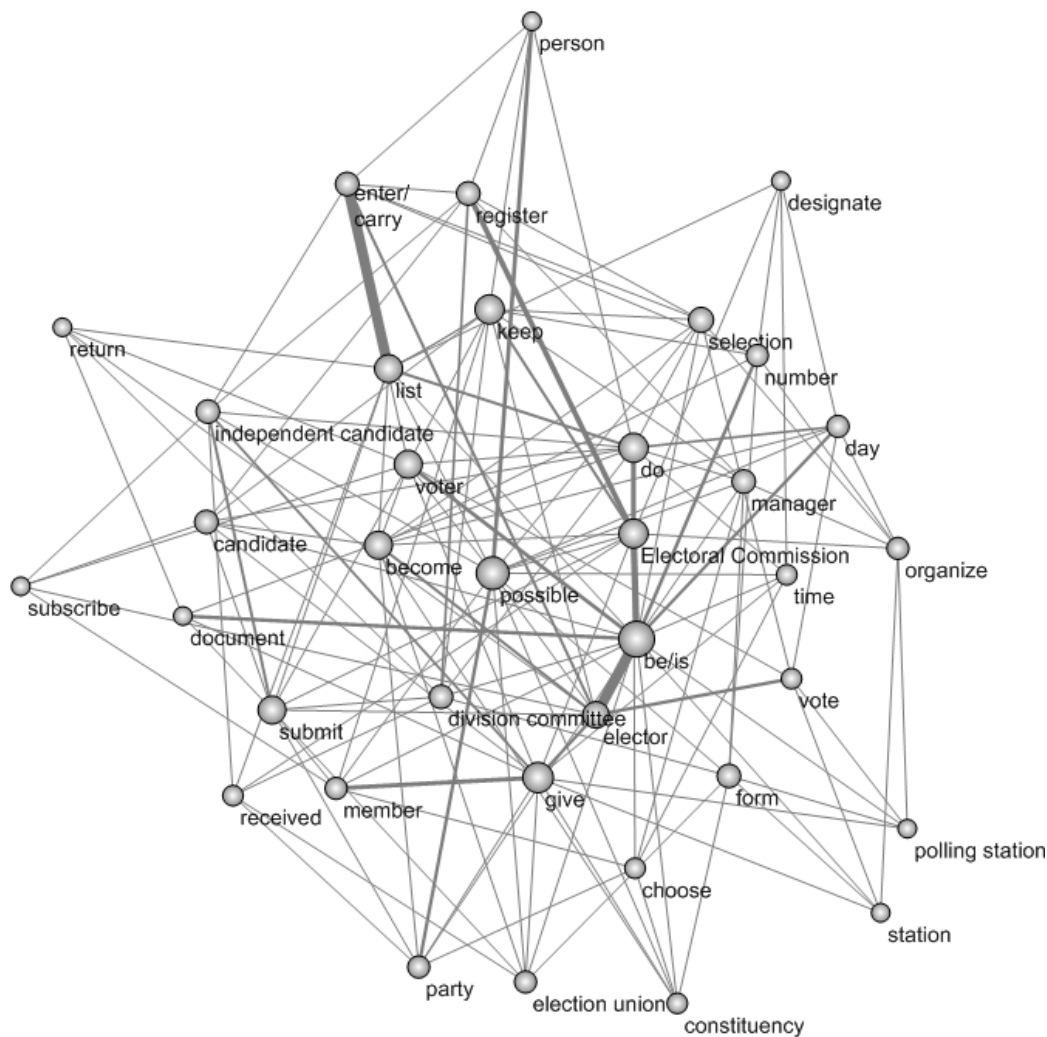
Table 2. The two-way table, showing the concordance of nouns and verbs within clauses (Täks, Kuusik, & Nyman-Metcalf, 2013)

Nouns\verb	be/i	submi	carry/ent	d	write	giv	kee	becom	organiz	thin
s	s	t	er	o	up	e	p	e	e	k

Electoral Commission	21	2	2	16	0	4	9	6	4	0
elector	34	3	5	0	9	10	3	9	0	0
list	6	3	25	10	8	0	2	3	0	0
appeal	4	20	0	0	0	0	0	0	0	0
person	9	11	14	3	4	4	5	0	0	0
day	10	0	5	8	0	0	0	2	1	27
party	10	8	12	0	0	4	0	2	0	0
division committee	3	4	0	4	0	0	0	0	0	0
member	4	7	7	0	0	14	4	0	0	0
candidate	4	1	0	1	0	2	2	2	0	0
voter	12	1	2	1	3	4	1	2	0	0
independent candidate	0	9	2	1	5	8	0	0	0	0
vote	8	0	0	1	0	0	2	0	24	0
envelope	0	0	0	0	0	0	0	0	0	0
decision	5	0	0	10	0	0	0	0	0	0
time	0	0	4	4	0	3	0	1	0	0
Family member	15	0	0	1	0	3	13	0	0	0

The number of words in this graph reached almost up to 600 words and the number of links was 1081. To make it easier to read the results, it is possible to filter out the most tense and almost fully connected network of keywords, a “skeleton” of the legal acts (*Figure 1*). The graph edges represent the words and the size of edges is showing the degree of it- to how many different words it connects. Arcs are showing the connections between words and the line thickness is showing how many times this pair of words did appear in one clause (max. 34 times).

Figure 1. Graph derived from verbs and nouns concordance in the clause



Extracting selected words from the text and preserving their connections for further analysis allows us to capture the norm content (60-80%), reduce the search base (more than 50%) and achieve qualitatively new level for automated analysis of legal texts. The two-way table forms a layer above the actual legal texts, reflecting the content of it at a very high level of abstraction, is machine readable and computable with help of different mathematical methods (data mining, graph computations etc.). Therefore such presentation of information is :

- *Scalable*- it can consolidate the norm, a subdivision, a chapter, legal act or whole legislation
- *Computable*- with help of math tools (data mining, graph theory etc.) it can be easily analysed by computers;
- *Visual*- it can be presented visually for human users and deliver high level overview about the content of the legal document;
- *Information enriched*- the picture of most frequent keywords (like Wordle), it is giving also some information about connections between different words and characterises

The visual representation of legal text as graphs also makes it possible to perform similarity measurements. Thus it is possible to find the use of the same nouns and verbs combinations (the “legal clause”) within different legal acts, and this way perform the quantitative analyses of the legal text, using different graph theory aspects and graph mining methods. The process of evaluating the similarity of two graphs is commonly referred to as graph matching, which aims to find a correspondence between the nodes and edges of two graphs that satisfies some more or less stringent constraints. In other words, similar substructures in one graph are mapped to similar structures in the other graph and extent of the result can be measured. (Aggarwal, Charu and Wang, 2010, p 219).

x.4.1 A test using Estonian legislation

In order to fit graph matching results for specific legislation analysis techniques and perform act to act similarity measurements, an experimental fitness function was generated. The function estimated the result in three stages- how many shared verbs were found, how many shared nouns connected to specific verbs and frequency of the use of each pair. The general weight is a sum of:

- Shared verbs account for 45%;
- Shared nouns connected to specific verbs 45%;
- Concordance frequency 10%.

Chosen particular calculation weights above are still under investigation in order to tune the measurement method, but results gained so far allows us to make some reasonably well justified conclusions already.

In order to test the method of normative system structural analysis:

- 386 Estonian legal acts were randomly chosen;
- Each legal act compared to each other legal act (148996 comparisons).

As a result a similarity table was created, consisting of 386 rows and 386 columns. Some interesting general characteristics appeared. In medium, all legal acts seem to share the content roughly by 1/3 but this rule has exceptions. For example in 167 cases (0,22%) two compared legal acts did not share any content. According to results it can be said that there is more likely overlapping between legal acts.

In 20 cases the similarity was measured 80% and higher. A quick control of two most similar legal acts (Estonian Parliament Election Act and European Parliament Election Act) showed a remarkable similarity of texts up to some parts exactly copied from each other.

In order to present a complex connected system, it has been useful to use extreme representation principles. In our case of Estonian legislation, the principle of maximal similarity spanning tree, successfully used for biological systems (Vohandu, 1961)., was used. For this principle, a similarity matrix S of legal acts, the maximal connected path was computed. In Figure 2 every legal act is connected to its most similar legal act. A zoomed upper part within rectangular area of this graph is shown in Figure 3.

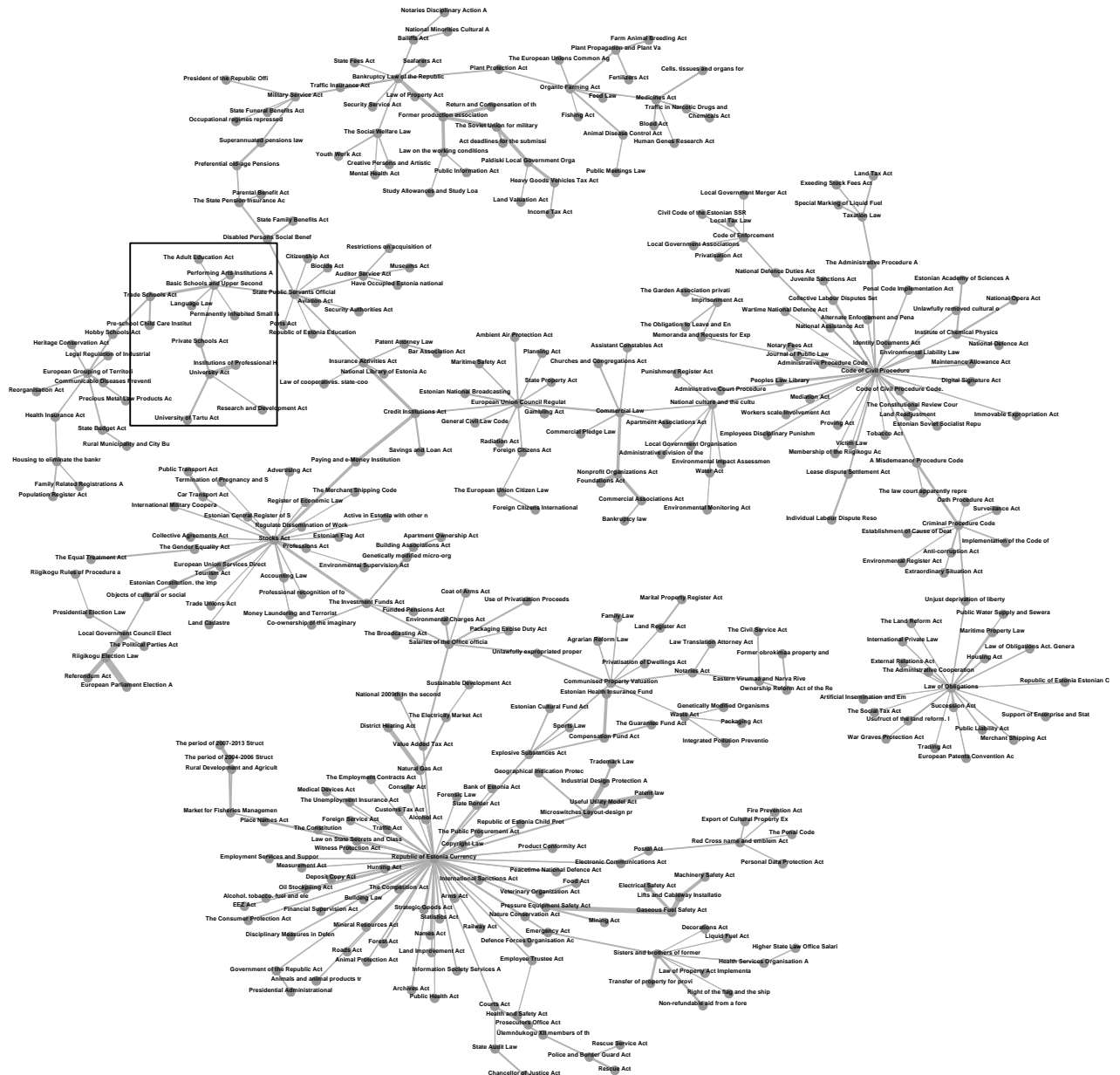


Figure 2. Similarity structure of the Estonian legislation, where each legal act is connected to the most similar one

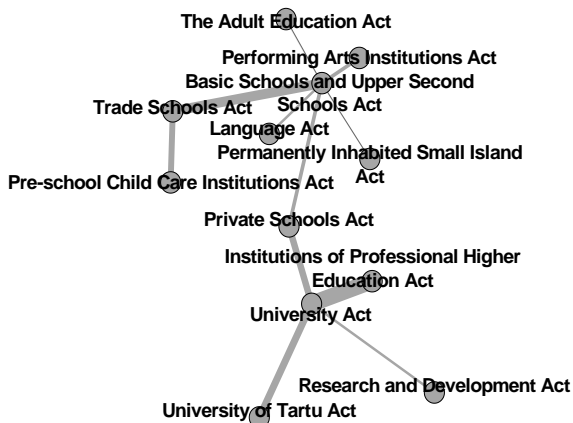


Figure 3. Extracted part of similarity structure

To match the clauses more precisely to each other and test the ability of the method for finding exact matches between legal sources originating from different locations.

4.2 Methodological refinement

In order to perform direct subgraph to subgraph comparisons and measure their similarity, we used the method described in the previous section and adjusted it according to the task. The developed method focused on norms (graph elements), originating from the clauses that are specific for the legal act, and compared these to each other graph element originating from other legal acts to pinpoint closely related norms within different texts and evaluate their similarity. The experimental fitness function described above was adjusted, and the concordance frequency was left out: the function estimated the result in two stages- how many shared verbs were found and how many shared nouns connected to specific verbs. The general similarity is a sum of:

- Shared verbs account for 50%;
- Shared nouns connected to specific verbs 50%;

Two legal acts (the Copyright Law, Law of Obligations Act) were chosen in order to test developed method. This choice was motivated by legal doctrinal research into copyright law reform. Kretschmer et al's (2010) analysis had shown often surprising and sometimes problematic dependencies between contract law and copyright law, and our approach could be one way to not only test these findings on a more abstract level for Estonian law, but also to utilize them strategically for law reform proposals. Short characteristics of the legal acts can be found in Table 3. After indexing, estimating the type and stemming the words found within legal case clauses (see Table 4) the second stage involved separating keywords from the clauses and creating a list of elementary graphs (groups of norm specific nouns and words together with identification information to relate them back to the legal act, see

Table 5). A similarity measurement according to the fitness function was performed in order to get actual similarity measurement results (see Table 6).

Table 3 Short characteristics of Copyright Law and Law of Obligations

Legal act	Clauses	Words	
the Copyright Law	818	13676	
Law of Obligations	5767	93770	

Table 4. Example of analysed legal text after indexing, estimating the type and stemming the words

Clause ID	Word ID	Words	Type	Basic form of word

97696	1231627	1	N	1
97696	1231628	kindlustusandja	S	kindlustusandja
97696	1231629	vabaneb	V	vabanema
97696	1231630	täitmise	S	täitmine
97696	1231631	kohustusest	S	kohustus
97696	1231632	kui	J	kui
97696	1231633	kindlustusvõtja	S	kindlustusvõtja
97696	1231634	kindlustatud	V	kindlustama
97696	1231635	või	J	või
97696	1231636	soodustatud	A	soodustatud
97696	1231637	isik	S	isik
97696	1231638	põhjustas	V	põhjustama
97696	1231639	kindlustusjuhtumi	S	kindlustusjuhtum
97696	1231640	toimumise	S	toimumine
97696	1231641	tahtlikult	D	tahtlikult

Table 5. Example of entry from the list of elementary graphs

Clause ID	Verb ID	Basic form of verb	Noun ID	Basic form of noun
§	2636	97696	97696	§
1231629	vabanema	1231628	kindlustusandja	1231629
1231634	kindlustama	1231630	täitmine	1231634
1231638	põhjustama	1231631	kohustus	1231638
		1231633	kindlustusvõtja	
		1231637	isik	
		1231639	kindlustusjuhtum	
		1231640	toimumine	

Table 6 Example of table of gained similarity measurement results

Legal act	Clause ID	Compared legal act:	Law of Obligations	Law of Obligations	Law of Obligations
the Copyright Law	59894	Compared legal act clause ID:	97697	97759	97836
		Similarity, %:	100	35	87,5

There were 32600 cases found where the similarity between two sub graphs was higher or equal to 30%; those were grouped based on similarity and given over to legal scientists in order to estimate the value of findings and usability of the results.

4.3 Legal Analysis

All 79 links that gave a 100% similarity were analysed. The purpose of the analyses was to understand whether the method links content which is related by some common features; or concepts that have a meaning in a legal discipline. Examples are discussed below.

Sample 1

Paragraph 14 of the Copyright Act regulates author's right to remuneration.⁷ The right to be paid for the use of author's work when author's economic rights are exercised is one of the main principles of copyright law. It is a general understanding supported by case law that whenever a work is used for business purposes either directly or indirectly, then author or a rightholder has to be compensated.⁸ A recording reproduced and sold is a form of a direct business, but a possibility of listening to a radio or watching a TV in a hotel room is a value adding element incorporated in the price of a room and therefore influences the booking of rooms indirectly. Copyright law may provide exceptions as to when it is not necessary to pay for the use of a work. Generally, when a person uses a work for private purposes, for instance, reproducing a CD for the purpose of sharing it with family and close friends, then it is not necessary to compensate the author. Exceptions derogating from the rule to pay are to be interpreted narrowly.⁹ The rule to pay the author is a fundamental principle in Copyright Law.

Subsection 3 of § 14 provides that it is prohibited to use a work before agreement is made between the author or the collective management organisation representing authors and a user of the work, specifying the amount of remuneration, and the collection and payment procedure.¹⁰ This subsection was linked to the second sentence „[a]ny agreement which

⁷ Copyright Act RT 1992, 49, 615; RT I, 28.12.2011, 1 (hereinafter Copyright Act), § 14.

⁸ See e. g. C-162/10 Phonographic Performance (Ireland) limited v. Ireland; C-306/05 Sociedad General de Autores y Editores de Espana (SGAE) versus Rafael Hoteles SA.

⁹ Exceptions have to conform to the three-step-test. Article 9 of the Berne Convention for the Protection of Literary and Artistic Works of 9 September 1886. Paris Act of 24 July 1971, as amended on 28 September 1979. WIPO, Geneva; § 17 of the Copyright Act.

¹⁰ Copyright Act § 14 (3).

derogates therefrom is void“ of § 452 (1); § 490 (2); § 677 (3); § 846 of the Law of Obligations Act.¹¹

As an example, subsection 1 of the paragraph 452 of the Law of Obligations Act provides that an insurer is not obliged to preform if the insured person intentionally caused the occurrence of the insured event.¹² It is followed by “*Any agreement which derogates therefrom is void*”.¹³ This is one of the fundamental principles of insurance law. The derogation is not possible because it would lead to the conflict of norms. If the derogation was possible, then it would mean that the insured person would be compensated for the commitment of crime, because the intentionally caused event can be the insurance fraud.¹⁴ The possibility of the conflict between the Law of Obligations and the Penal Law is excluded by the imperative norm stating that it is not possible to make an agreement contrary to this principle.

Similarly, if the work is used without the permission of the author, then it can be a criminal act resorting to the Penal Law.¹⁵ It is imperative that the agreement has to be reached.

Sample 2

Copyright law regulates the ownership of Copyright protected works created under the employment relationship. In case the employment contract is silent on the transfer of author’s economic rights, then the transfer to the employer is presumed as the general principle of Copyright law. It follows the idea that the investor is entitled to the fruits of the investment. This is the case in Estonia. The purpose of this rule is to provide legal certainty, because copyright provisions are absent in most employment contracts. Yet, an employer and employee are always free to regulate the transfer of Copyright differently from the default rule.

Similarly, the producer becomes the owner of author’s economic rights in the audio-visual work, because the transfer of rights is presumed by law. The transfer of rights is often a practicality necessary for the investor to be able to manage the business related to the work.

¹¹ Law of Obligations Act RT I 2001, 81, 487; RT I, 05.04.2013, 1 (hereinafter Law of Obligations Act), §§ 452 (1); 490 (2); 677 (3); 846.

¹² Ibid. § 452 (1).

¹³ Ibid.

¹⁴ Penal Code RT I 2001,61,364; RT I 29.12.2011,1, § 212.

„§ 212. Insurance fraud

(1) A person who intentionally brings about an insured event or causes a misconception of the occurrence of an insured event with the intention to receive an insurance indemnity from the insurer shall be punished by a pecuniary punishment or up to 5 years’ imprisonment.

(2) The same act, if committed by a legal person, is punishable by a pecuniary punishment.“

¹⁵ Ibid. § 223.

„§ 223. Unlawful direction of works or objects of related rights towards public

(1) Unlawful public performance, showing, transmission, re-transmission or making available to the public or a work or an object of related rights for commercial purposes is punishable by a pecuniary punishment or up to one year of imprisonment.

(2) The same act, if performed by using a pirated copy, is punishable by a pecuniary punishment or up to 3 years’ imprisonment.

(3) An act provided for in subsection (1) or (2) of this section, if committed by a legal person, is punishable by a pecuniary punishment.

(4) The court shall confiscate the object which was the direct object of the offence provided for in subsection (2) of this section.“

Subsection 6 of § 14 of the Copyright Act regulates the author's right to remuneration after the transfer of economic rights has been agreed by contract or presumed by law.¹⁶ According to this subsection the producer of the audio-visual work may become the sole owner of the economic rights, but the author still retains the right to equitable remuneration. The producer has the sole right to decide about the reproduction of the work, the adaptation of the work, and the way it is made available to the public. But if proceeds are made as a result of these decisions, for instance, making a movie available through a broadcasting organisation, then the author is entitled to receive the equitable share from these proceeds.

Subsection 7 of § 14 of the Copyright Act provides a similar regulation concerning the commercial rental of the phonogram.¹⁷ Author may agree with the phonogram producer on the transfer of the economic rights or it is presumed by law, then the author is entitled to the equitable share of the proceeds received from the rental of the phonogram.

Both subsections provide that “[a]n agreement to waive the right to obtain equitable remuneration is void”.¹⁸ These subsections were connected to the provisions of § 452 (1); § 490 (2); § 672; § 677 (3); § 846 of the Law of Obligations Act by the sentence „Any agreement which derogates therefrom is void”.¹⁹

Sample 3

Furthermore, according to § 67 (1) of the Copyright Act „[a] performer has the exclusive right to use and to authorise or prohibit the use of the performance of a work and to obtain, for such use, remuneration agreed upon by the parties except in the cases prescribed by this Act and an agreement between the parties”.²⁰ This subsection was linked to the provisions of § 452 (1); § 490 (2); § 677 (3); § 846 of the Law of Obligations Act by the sentence „Any agreement which derogates therefrom is void”.²¹ It is imperative that the agreement has to be reached on how exclusive rights are used and remunerated under Copyright law, because the breach of these principles is criminalised.

Sample 4

§ 32 (1) of the Copyright Act provides that author shall enjoy Copyright in works created under an employment contract or in the public service, but the economic rights shall be transferred to the employer unless otherwise agreed by contract.²² This subsection is linked to § 8 (2) of the Law of Obligations, providing that “[a] contract is binding on parties”.²³

Although the norm is dispositional, the presumption of transfer of economic rights prescribed in Copyright law is an important exception, because it is not evident that on the basis of general principles of contract law parties should transfer rights automatically without any reflection on the intent to do so. *Pacta sunt servanda* is the moral imperative in contract law, but as a norm it is dispositional in a sense that parties are free to decide upon conditions of withdrawal from a contract. Both principles allow derogations provided that parties prescribe

¹⁶ Copyright Act § 14 (6).

¹⁷ Ibid. § 14 (7).

¹⁸ Ibid.

¹⁹ Law of Obligations Act, §§ 452 (1); 490 (2); 672; 677 (3); 846.

²⁰ Copyright Act, § 67 (1).

²¹ Law of Obligations Act, §§ 452 (1); 490 (2); 677 (3); 846.

²² Copyright Act, § 32 (1).

²³ Law of Obligations Act, § 8 (2).

the transfer of author's economic rights in a contract differently from the default rule of the Copyright law, and on the bases of contract law, parties agree on the conditions of withdrawal in a contract, but until they have not done so, the link between these two principles suggests that the automatic transfer of economic rights is binding.

5 Conclusion and future work

The links revealed by this method suggest that it is possible to pinpoint imperative norms and some core principles important to the Copyright law and the Law of Obligations, which are not "obvious" or linked through explicit citations. Many of these links do not have similarity based on the content of the norms presiding the imperative clause. For instance, provisions of insurance law in general have no influence on how the author's right to remuneration has to be understood in the context of Copyright law and general principles of contract law. However, the value of this method resides in the fact that it enables to identify imperative norms scattered in different legal acts of private law. The possibility to identify imperative norms as oppose to dispositional norms would be an effective tool in itself, because it would enable a systemized identification of the critical elements of law in the process of amending legal acts, or it gives an insight to any stakeholder who needs to retrieve information about the norms which limit the contractual freedom.

The experiment has given preliminary results, which have to be further analysed. It is an open question for us to what extent this method identified imperative norms contained in the Copyright Act and the Law of Obligations Act, because only links of 100% similarity were analysed, and there is no research material indicating the number of imperative norms contained in those legal acts. Therefore no definite result, which could be calculated based on the comparison, can at this point be stated. It can be said that 82,3% links out of 79 links that rendered 100 % similarity have a connection which has a meaning in the legal study. However, as a proof of concept the analysis fulfills the two tasks that we set ourselves in the introduction. It allows the legislator to identify "hidden" connections between laws. Taking Wintgens' Principle of Coherence as a starting point and putting the notion of a legal system at the heart of our analysis, the analysis can then alert the legislator on the one hand to unforeseen consequences of legislative reform, for instance if a change in definition in one part of the system can also impact on word meaning in a superficially unrelated part. Conversely, these hidden connections can work as a heuristic to improve PA, and identify non-obvious alternatives for legislative reform that may be less intrusive on citizen's freedom, e.g. by adding enabling norms in contract law, we may avoid sanction introducing norms in copyright law. Finally, for law firms it allows to estimate just how much their business will be affected by a proposed law reform, as it shows "at a glance" the dependency pathways, the way the specific reform can impact on the legal system as a whole. In the future, apart from evaluating larger segments of Estonian regulation, we hope to combine this method with citation analysis tools, to see in more detail how explicit and implicit interconnection between norms can be data mined to assist the legislative process, and to bring our proxies closer to the semantic content of Wintgens' four principles of rational jurisprudence. The result, we hope, will be a practical computational tool that assists legislators and addresses of legislation, while being at the same time deeply rooted in jurisprudential analysis.

6 REFERENCES

Aggarwal, Charu C., and Haixun Wang, 2010, *Managing and mining graph data*. Vol. 40. NY: Springer

Aitken, Gary, 1997, The law of BSE. *Journal of the Law Society of Scotland* 42 17 – 20

Apistola, M., & Lodder, A. R. 2005. Law firms and IT. Towards optimal knowledge management. *Journal of Information, Law and Technology*, 2(3), 1–28

Ashtar, Reuven. 2010 "Licensing as Digital Rights Management, from the Advent of the Web to the iPad." *Yale JL & Tech.* 13 p: 141

Barry, Nicholas, 2012-2013, Man versus Machine Review: The Showdown between Hordes of Discovery Lawyers and a Computer-Utilizing Predictive-Coding Technology; 15 *Vand. J. Ent. & Tech. L.* 343

Baron, Jason R. 2011 "Law in the Age of Exabytes: Some Further Thoughts on 'Information Inflation' and Current Issues in E-Discovery Search." *Rich. JL & tech.* 17 p: 9-16

Becchetti, Leonardo, David Andres, Londono Bedoya, and Paganetto, Luigi. 2003, ICT investment, productivity and efficiency: evidence at firm level using a stochastic frontier approach, *Journal of productivity analysis* 20.2 143-167.

Becker, E. (Ed.). (2003). *Digital rights management: technological, economic, legal and political aspects* (Vol. 2770). Springer

Beger, Gabriele, 2010, Open Access und die dritte Reform des Urheberrechts in der Informationsgesellschaft. *PIK-Praxis der Informationsverarbeitung und Kommunikation* 33.3 185-186.

Bench-Capon, T. J., & Coenen, F. P. (1992). Isomorphism and legal knowledge based systems. *Artificial Intelligence and Law*, 1(1), 65-86.

Berman, Donald H., & Hafner, Carol, 1989, The potential of artificial intelligence to help solve the crisis in our legal system. *Communications of the ACM*, 32(8), 928-938

Blichner, Lars Chr, and Anders Molander, 2008, Mapping juridification, *European Law Journal* 14.1: 36-54

Bommarito II, Michael J., and Daniel M. Katz, 2010, A mathematical approach to the study of the united states code, *Physica A: Statistical Mechanics and its Applications* 389.19: 4195-4200

Boose, John H., 1989, A survey of knowledge acquisition techniques and tools." *Knowledge acquisition* 1.1: 3-37

Boulet, Romain, Mazzega, Pierre and Bourcier, Danièle, 2011, A network approach to the French system of legal codes—part I: analysis of a dense network. *Artificial Intelligence and Law* 19.4: 333-355

Bourcier, D, Casanovas P, Dulong de Rosnay, M, Maracke C (2010) (Eds.) *Intelligent Multimedia: Managing Creative Work in a Digital World*, European Press Academic Publishing

Brazier, Margaret. 1997 Hard cases make bad law? *Journal of medical ethics* 23. 6 341- 343

Breuker, Joost, Petkov, Emil, and Winkels, Radboud, 2000, Drafting and validating regulations: the inevitable use of intelligent tools." *Artificial Intelligence: Methodology, Systems, and Applications*. 21-33 Heidelberg: Springer

Chikofsky, Elliot J., and Cross James H., 1990, Reverse engineering and design recovery: A taxonomy." *Software, IEEE* 7,1: 13-17

Cohen, Julius, 1949-1950, *Towards Realism in Legisprudence*, 59 Yale L. J. 886-897

Conrad, Jack G. 2010 "E-Discovery revisited: the need for artificial intelligence beyond information retrieval." *Artificial Intelligence and Law* 18, no. 4 p: 321-345.

Contissa, G., & Laukyte, M. 2008. Legal Knowledge Representation: A Twofold Experience in the Domain of Intellectual Property Law. In *JURIX* (pp. 98-107)

Davis, Lewis Anthony, 1981, Review Procedures and Public Accountability in Sunset Legislation: An Analysis and Proposal for Reform." *Administrative Law Review* 393-413

Favale, M. 2011. Approximation and DRM: can digital locks respect copyright exceptions? *International Journal of Law and Information Technology*, 19(4), 306-323

Finn, John E., 2009, Sunset clauses and democratic deliberation: assessing the significance of sunset provisions in antiterrorism legislation." *Colum. J. Transnat'l L.* 48: 442.

Forester, Tom, (ed). 1985, *The information technology revolution*. Cambridge (Mas): MIT Press

Fuller, Lon L. 1969, *The morality of law*. Yale: Yale University Press,

García, R, Gil, R and Delgado J (2007). "A web ontologies framework for digital rights management." *Artificial Intelligence and Law* 15.2 137-154.

Hafner, Carole D., and Marc Lauritsen, 2007, Extending the power of automated legal drafting technology. *Proceedings of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference*, 59-68 Amsterdam:IOS Press

Hargeraves, Ian, 2011, *Digital Opportunity: A review of intellectual property and growth*, available at <http://www.ipo.gov.uk/ipreview-finalreport.pdf>;

Hart, Herbert Lionel Adolphus, 1957, Positivism and the Separation of Law and Morals, *Harv. L. Rev.* 71: 615-621

Hoekstra, Alexander Boer Radboud Winkels Rinke, and Tom M. van Engers, 2003 Knowledge management for legislative drafting in an international setting. *Legal Knowledge and Information Systems: Jurix 2003: the Sixteenth Annual Conference*. 91-100 Amsterdam; IOS Press

Hogan, Christopher, Robert S. Bauer, and Dan Brassil. 2010 "Automation of legal sensemaking in e-discovery." *Artificial Intelligence and Law* 18.4 p 431-457.

Hutchins, Christopher E., and Marge Benham-Hutchins, 2010, Hiding in plain sight: criminal network analysis, *Computational and Mathematical Organization Theory* 16.1: 89-111

Kim, Rakhyun E. 2013, The emergent network structure of the multilateral environmental agreement system. *Global Environmental Change* 23.5: 980-991

Kretschmer, M., Derclaye, E., Favale, M. and Watt, R. (2010), The Relationship between Copyright and Contract Law: A Review commissioned by the UK Strategic Advisory Board for Intellectual Property Policy (SABIP) (172pp), available at: <http://www.ipo.gov.uk/ipresearch>

Kretschmer, Martin and Towse, Ruth, (eds), 2013, What Constitutes Evidence for Copyright Policy? Digital proceedings of ESRC symposium www.copyrightevidence.org/create/esrc_evidence_symposium

Kretschmer, Martin, 2008, Creativity stifled? A joint academic statement on the proposed copyright term extension for sound recordings. *European Intellectual Property Review*, 30.9: 341-374

Kysar, Rebecca, 2011, Lasting Legislation. *University of Pennsylvania Law Review* 159 (2011): 101

Lessig, Lawrence, 1999, *Code and other laws of cyberspace*, NY:Basic Books

McCarty, L. Thorne, 1980, The TAXMAN project: Towards a cognitive theory of legal argument, *Computer Science and Law: An Advanced Course*, 23-43

Mendis, D., and M. Kretschmer, 2013, *The Treatment of Parodies Under Copyright Law in Seven Jurisdictions: A Comparative Review of the Underlying Principles*, Intellectual Property Office available at <http://www.ipo.gov.uk/ipresearch-parody-report2-150313.pdf>

Oskamp, A., & Lauritsen, M. 2002. AI in Law Practice? So far, not much. *AI and Law*, 10(4), 227-236

O'Leary, Daniel E. 1998, Enterprise knowledge management, *Computer* 31.3: 54-61

- Palmirani, Monica, 2010, Legislative change management with Akoma-Ntoso." *Legislative XML for the Semantic Web*. Springer Netherlands, 101-130
- Patry, William, 2012, *How to fix copyright*, N.Y.: Oxford University Press
- Prandoni, C., Valentini, M., & Doerr, M. (2009). Formalising a model for digital rights clearance. In *Research and Advanced Technology for Digital Libraries* (pp. 327-338). Springer Berlin Heidelberg
- Samuelson, Pamela, 2013, Is copyright reform possible?, *Harv. L. Rev.* 126 740-779.
- Schwartz, Daniel M., and Tony DA Rouselle. 2009, Using social network analysis to target criminal networks, *Trends in Organized Crime* 12.2: 188-207
- Sergot, Marek J., et al. 1986, The British Nationality Act as a logic program. *Communications of the ACM* 29.5: 370-386
- Sommerville, Ian, and Gerald Kotonya, 1998, *Requirements engineering: processes and techniques*. Hoboken:John Wiley & Sons
- Staudt, Ronald W. 2003 "Perspectives on knowledge management in law firms." LexisNexis
- Staudt, Ronald W., 2008, All the Wild Possibilities: Technology That Attacks Barriers to Access to Justice. *Loy. LAL Rev.*, 42: 1117-1146
- Swierczek, Fredric William, Pritam K. Shrestha, and Clemens Bechter, 2005, Information Technology, Productivity and Profitability in Asia-Pacific Banks, *Journal of Global Information Technology Management* 8.1: 6-26
- Täks, Ermo; Kuusik, Rein; Nyman-Metcalf, Katrin Merike, 2013, An automated legal content capture and visualisation method Tallinn: Tallinn University Press
- Tehrani, John, 2007, Infringement nation: Copyright reform and the law/norm gap." *Utah L. Rev.* 3:537-550
- Towse, Ruth, 2011, What we know, what we don't know and what policy-makers would like us to know about the economics of copyright', *Review of Economic Research on Copyright Issues*, 8:101-120
- Voermans, Wim JM, Remco Van Wijk, and Fokkema, Welmoed, 2012, Free the Legislative Process of its Paper Chains: IT-inspired Redesign of the Legislative Procedure In Snellen, I.Th.M., Thaens, M., Van De Donk (eds) , *Public Administration in the Information Age: Revisited* Amsterdam :IOS Press 237 – 251
- Voermans, Wim, Luuk Matthijssen, and Egon Verharen, 1997, Legimatics in Europe: the LEDA-project, *New review of applied expert systems* 3: 155-167
- Vohandu, Leo. K., 1961, Analysis of Biological Systems with many Variables." In *III Conference of Mathematical Methods in Biology*, 19-22.

Winkels, Radboud, and Nienke Den Haan, 1995, Automated legislative drafting: Generating paraphrases of legislation." In *Proceedings of the 5th international conference on Artificial intelligence and law*, 112-118. New York: ACM

Wintgens, Luc J., 2006, Legisprudence as a new theory of legislation." *Ratio Juris* 19.1: 1-25.

Xu, Jennifer J., and Hsinchun Chen, 2005, CrimeNet explorer: a framework for criminal network knowledge discovery. *ACM Transactions on Information Systems (TOIS)* 23.2 201-226.

Zhang, X. M., Liu, Q., & Wang, H. Q. 2012. Ontologies for intellectual property rights protection. *Expert Systems with Applications*, 39(1), 1388-1400



RCUK Centre for Copyright and
New Business Models in the
Creative Economy

College of Social Sciences / School of Law
University of Glasgow
10 The Square
Glasgow G12 8QQ
Web: www.create.ac.uk

