THE ROLE OF AI & LAW IN LEGAL DATA SCIENCE

Erich Schweighofer

Professor, University of Vienna
Centre for Computers and Law, Faculty of Law
Schottenbastei 10-16/2/5, 1010 Vienna, AT
Erich.Schweighofer@univie.ac.at, http://rechtsinformatik.univie.ac.at

Abstract: *The potential of AI & law methods in law has been not properly used. A promising way out may be legal data analysis. The goal of legal data science is to complement the existing methodology of law with the new computer-based methods, and to bring it into a theoretical framework. In previous research, we have developed the 8 views/4 methods/4 syntheses approach of legal data analysis. In this paper, we will focus on the man/machine delivery of the desired products of legal knowledge representation using AI & law methods. At present, a lot of this analysis is done manually but the lack of sufficient resources becomes more and more evident. So far, tools of data analysis are insufficiently developed and used in the legal domain. This paper should describe the potential of this approach in order to motivate a stronger deployment in the analysis of legal text corpora.*

1. Introduction

The potential of AI & law methods in law has been mostly not properly used. Thus, too many traditional lawyers consider this branch of legal science as irrelevant, at least for the time being. A promising way out may be a highly developed and semi-automatic analysis of the legal system using a better systematisation of knowledge and analytical tools for lawyering. The goal of legal data science is to complement the existing methodology of law with the new computer-based methods, and to bring it into a theoretical framework. In previous research, we have developed the 8 views/4 methods/4 syntheses approach of legal data analysis.[1] The need is evident, considering that a textual representation is insufficient for today's needs of the legal profession. Paraphrasing the words of LAYMAN ALLEN in 1957[2], legal data science provides razor sharp tools for the analysis of the legal system. «The fight for justice and law» – in the sense of RUDOLF VON JHERING[3] – goes on, but now including legal data science.

In this paper, we will focus on the delivery of the desired products of legal knowledge representation using also AI & law methods. The options of knowledge can be systemized as follows: Besides the text (multimedia) corpus, 7 other views are now – more or less – required: metadata, citation network, user view, logic view, ontological view, visualisation view and argumentation view. Further, lawyers like a synthesis of the views in a practice-oriented form: commentary or manual, Dynamic Electronic Legal Commentary (DynELC)[4], citizen

---

[1] E. SCHWEIGHOFER, From Information Retrieval and Artificial Intelligence to Legal Data Science. In *Proceedings of the ICAIL MWAIL2015 Workshop*, pages 13–24, OCG, Vienna 2015.

[2] L. ALLEN, Symbolic Logic: A Razor-Edged Tool for Drafting and Interpreting Legal Documents. In *The Yale Law Journal* 66, 833–879, 1957.

[3] R. v. JEHRING, Der Kampf ums Recht, Vortrag (The fight for law), Vienna 1872. Schutterwald/Baden 1997, 1872.

[4] E. SCHWEIGHOFER, Indexing as an ontological-based support for legal reasoning. In: J. Yearwood and A. Stranieri (eds.), *Technologies for Supporting Reasoning Communities and Collaborative Decision Making: Cooperative Approaches*, pages 213–236, IGI Global Publishers, Hershey 2011.

information system and case-based synthesis. A lot of this analysis is done manually but the lack of sufficient resources becomes more and more evident. A strong co-operation between man and machine is required.

Legal information retrieval has been around for a long time but the analysis has to be improved in order to allow higher recall and better ranking. A semi-automatic analysis for metadata, citations and ontologies can be obtained by computational linguistics and machine learning. Further information can be derived with document categorization and the semi-automatic generation of thesaurus descriptors; also considering multilingual jurisdictions. The tricky but manageable automatic generation of hypertext links should be supplemented by an efficient ranking. The automatic generation of temporal relationships allows the production of consolidated versions. Data-friendly generation of user behaviour gives strong insights of their assessment. The semi-automatic generation of summaries of documents allows a much easier and more efficient analysis of documents. Highly needed tools for segmentation of texts into logical rules and arguments exist but have to be improved. Visualization tools allow an easier understanding of complex rules. The user needs are taken into account in sophisticated ranking algorithms: document vs. requirements, document in the corpus, document in the citations network, document in the timeline etc.

Legal data analysis is still insufficiently developed. Experts do not use such tools in order to generate synthesis tools like commentaries. This paper describes the potential, the state of the art and shows existing strengths and lacunae.

2. Legal system & textual representation

Any comprehensive theory of legal data science requires a definition of the law as a textual or multimedia representation. With HANS KELSEN[5] the law is seen as a normative order that governs the behaviour of people. A norm belongs to this system when «man should behave in certain ways because of the norm.» In cyberspace with its multilevel regulatory structure, an extension is needed. Law is understood as the basic consensus of the stakeholders of the legal systems on «ought». Legal informatics describes this body of legal rules, e.g. the legislative processes and the respective documents. These documents are the core of the legal retrieval system (legal information system). This is still primarily done in a textual way, but is increasingly being enriched with multimedia elements. This corpus is huge: at least several gigabytes of data, millions of documents over a million records in the database dictionary, etc. Recently, an extension of legal materials to implementation practice or social web data can be observed or is strongly considered.

3. Goals of legal data science: creating 8 views of the legal system

The goal of legal data science is to complement the existing methodology of law with the new computer-based methods, and to bring it into a theoretical framework. For a long time, work on legal knowledge has been only in the focus of legal theory and information science. Only since the late 1950s – with the start of research on legal information retrieval – appropriate progress can be noted (term retrieval, text retrieval, metadata, citations, search technologies, user interface, telecommunications, etc.). The theory distinguishes 8 views – text view, metadata view, network view, user view, logic view, ontological view, visualisation view and argumentation view.

In the last decades, text corpora with search engines have become the standard of knowledge representation but are now considered as insufficient. The «views theory» of LU and CONRAD[6] – documents view, annotation view, citation view and user view describes this enlargement and constitute a basis for this work. However, it

---

[5] H. KELSEN, Reine Rechtslehre (Pure Theory of Law), 2nd ed., pages 196 et seq., 1960.

[6] Q. LU, J. CONRAD, Next generation legal search – it's already there. In: *Cornell Legal Information Institute*, VoxPopuLII, https://blog.law.cornell.edu/voxpop/2013/03/28/next-generation-legal-search-its-already-here/, 2013 (last accessed: 12 January 2016).

should be extended by four other views – legal logic, legal ontologies, legal visualisation and legal argumentation. Further, the works of SOWA[7], FIEDLER[8], ZELEZNIKOW/HUNTER[9], the authors of the book collection of YEARWOOD/STRANIERI[10] as well as our own work on knowledge representation of the law[11] have been taken into account.

The formalization of the law has started in the 1970s; the research is well documented in the proceedings of the conferences ICAIL[12] and JURIX[13]. In this paper, we argue for a stronger integration of methods of information retrieval, AI & law, legal language processing but also using all methods of data analysis useful for the legal discipline.

Knowledge representation of law is not just about the documentation itself; each view represents further insights on the law itself. The use of all available media and methods – language, meta-knowledge, visualisation, structure, mathematics and statistics, logic, ontologies, formal structures, etc. – implies also that legal «knowledge sources» must be integrated into this knowledge structure, however, keeping the holistic and harmonious picture. Language, visualisation, structure, etc. are also essential methods of human thinking itself. The principle applies that more quantity of methods and analysis ultimately results in higher quality. The more comprehensively all media, factors and methods are used, the better is the structural analysis of the law.

The text corpus (LU/CONRAD: textual set of evidence) consists of all relevant documents of the legal system: statutes, regulations, court decisions, literature, administrative practice, as well as complementary legal practice such as briefs or draft legislation. The basis is the triad of norms, court decisions and literature, however, to be extended by the now huge body of «soft law». Nowadays, there are hundreds of types of documents in any jurisdiction. Recently, implementation practice and social networks are getting higher relevance.

The description view contains the metadata of the law (LU/CONRAD: annotation view). Information science has developed a good methodology to describe materials in order to classify and summarize important content or extract relevant information. The quality of this description is highly variable (e.g. the high-level West's Key Number System[14] or the formidable CELEX metadata system[15]). In general, metadata is only useful for those with sufficient familiarity. However, AI & law methods allow intelligent use of metadata, e.g. setting the proper context of the communication, e.g. persons, roles, purpose, topic etc. without much user input.

The citation network *(*LU/CONRAD: citation network view) describes the relations between the documents. It has always been a key topic of legal documentation and is and will remain indispensable. Formally, it is documented if a document cites others ((out-bound (cited) sources) and if it is quoted by others (in-bound (citing) sources)). It is important to specify citations, e.g. referring also to the structuring elements of a

---

[7]   J. F. SOWA, Knowledge representation: logical, philosophical, and computational foundations. Course Technology, Boston 2000.

[8]   H. FIEDLER, Modell und Modellbildung als Themen der juristischen Methodenlehre (Model and modeling as subjects of legal methodology). In: *Proceedings of the International Legal Informatics Conference IRIS2006,* pages 275–281, OCG, Vienna 2006.

[9]   J. ZELEZNIKOW, D. HUNTER, Building Intelligent Legal Information Systems, Representation and Reasoning in Law, Computer Law Series 13, Kluwer, Deventer 1994.

[10]   J. Yearwood and A. Stranieri (eds.), *Technologies for Supporting Reasoning Communities and Collaborative Decision Making: Cooperative Approaches*, IGI Global Publishers, Hershey 2011.

[11]   See footnote 4.

[12]   The ICAIL proceedings are published with ACM Publishers, New York and available in the ACM Digital Library. Latest edition: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, Rome 2015: http://dl.acm.org/citation.cfm?id=2514601 (last accessed: 12 January 2016).

[13]   Website listing JURIX proceedings: http://jurix.nl/proceedings/ (last accessed: 12 January 2016).

[14]   Westlaw, West Key Number System® on WestlawNext®: thomsonreuters.com/pdf/wln2/L-374484.pdf (last accessed: 12 September 2015).

[15]   E. SCHWEIGHOFER, Wissensrepräsentation in Information Retrieval-Systemen am Beispiel des EU-Rechts, Dissertation, Universität Wien 1995, (knowledge representation in information retrieval systems on the example of EU law, PhD thesis, University of Vienna 1995, published in extended version WUV publishers, Vienna 2000).

document, e.g. articles, sections, paragraphs, lists, etc. The main types of citations are: basis of the act, cited acts in the document, citations in the operative part of the judgment, document amending other documents, document is amended by other acts, etc.[16].

The user's perspective takes into account the opinions of legal professions, business and civil society about the document collection *(*Lu/Conrad: user view*)*. Modern search engines can (semi)automatically generate user's perspectives. The search engine can collect information how often a particular document has been opened for viewing, how many times it has been printed, how many times it has been tested for its legal validity, etc. Thus, information about the document relevancy is available that is otherwise difficult to obtain. Sufficient data protection has to be implemented, e.g. anonymization of user data.[17]

The logical representation is based on the first order logic, complemented by a representation of temporal layers[18] and the personal scope[19]. The decisive factor is the rapid and efficient execution of an at least five-digit number of rules (i.e. more than 10,000 rules). These requirements have been met in practice but are still insufficiently used.[20] Logic programming restricts the complexity of the logic view but allows the handling of «standard cases» in high volumes at low cost. Deontic logic tries to improve the theoretical basis for advanced logical representations, e.g. moving to solve also «hard cases». So far, only a logical representation of argumentation structures is considered as possible.

The conceptual representation can draw on the experience of «concept jurisprudence» *(Begriffsjurisprudenz)* that has intensively dealt with the conceptualization and systematisation and was dominant in the 19th century. The conceptual jurisprudence is a tool of thought for a better understanding of the legal system. The legal ontology, however, brings the automation option and a much more powerful systemisation. Since the 1990s, many legal ontologies have been developed[21]; now a strong standard exists with LRI Core, LKIF[22] and LegalRuleML[23]. Much research has been reused for a more convenient representation of legal knowledge[24], [25]. The respective elements of the concepts have to be transposed into a computer-readable structure, e.g. header, definition, relations (upper/lower term, antonym, related term, synonym, homonym, polysem etc.), pre-subsumption (relation between the concept of law and facts element) and comments. Facts, e.g. a world ontology, can be taken from projects like Cyc. This project is to provide automated applications with a formal knowledge base of «common sense» knowledge. Currently, more than three million facts and rules are for-

---

[16] A. Berger, The development of references in the legislative documentation, Verlag Dokumentation, Pullach near Munich 1971.

[17] J. Conrad, presentation (Skype) at the OCG digital2014 Conference in November 2014 in Vienna.

[18] J. Scharf, Wissensrepräsentation und automatisierte Entscheidungsfindung am Beispiel des KOVG (Knowledge representation and automated decision making), PhD thesis, University of Vienna 2015.

[19] Idea of Erich Schweighofer, based on relevant thoughts of C. Reed, You Talkin' to Me?. In: Jon Bing, *en hyllest, a tribute*, pages 154–171, 2014.

[20] A prime example is the Australian company SoftLaw; this was subsequently acquired by Oracle; the application itself is available as Oracle Business Rules.

[21] G. Sator, P. Casanovas, M. A. Biasiotti, M. Fernández-Barrera, Approaches to Legal Ontologies: Theories, Domains, Methodologies, Springer, Dordrecht/Heidelberg/London/New York 2011.

[22] R. Hoekstra, J. Breuker, M. De Bello, A. Boer, The LKIF Core Ontology of Basic Legal Concepts. In: P. Casanovas, M.A. Biasiotti, E. Francesconi, M.T. Sagri (eds.) *Proceedings of LOAIT 07, II. Workshop on Legal Ontologies and Artificial Intelligence Techniques,* pages 43–64:

[23] Website OASIS LegalRuleML, https://www.oasis-open.org/committees/ tc_home.php?wg_abbrev=legalruleml (lastly accessed: 12 January 2016).

[24] N. Casellas., E. Francesconi, R. Hoekstra, S. Montemagni (eds.), Proceedings of LOAIT 2009, 3rd Workshop on Legal Ontologies and Artificial Intelligence Techniques joint with 2nd Workshop on Semantic Processing of Legal Text. IOT Series, Barcelona 2009.

[25] P. Casanovas, M.A. Biasiotti, E. Francesconi, M.T. Sagri (eds.) *Proceedings of LOAIT 07, II. Workshop on Legal Ontologies and Artificial Intelligence Techniques*: http://www.ittig.cnr.it/loait/LOAIT07-Proceedings.pdf, 2007.

mally represented in the Cyc knowledge base, provided as OpenCyc.[26] A major advantage of an ontology is the easier representation of relations between facts and legal concepts in a form of pre-subsumption.

The theory development of legal ontologies and also its practical implementation are sufficiently strong, however, in practice, few applications exist. Legal ontologies require a deep analysis of the legal domain that can only be done by the legal expert. So far, very few experts have moved into this new form of representation of knowledge. Thus, legal ontologies share the fate of the Semantic Web[27]. Semantic mark-up is a strong tool for a co-operative analysis of legal text corpora but requires a more efficient collaboration between knowledge experts and knowledge engineers that rarely exists.

Following COLETTE R. BRUNSCHWIG, legal visualization concerns the use of graphics, images and videos for visual representation of the law.[28] Visualisation draws attention to the essentials, makes these memorable, increases understanding and shows hidden connections. The potential of visualizations for dialogue with the laity is undisputed. Graphical notations are also a support for the formalization of the law. The key features are represented by images or graphics, even in cases where the necessary level of abstraction for formalization is not yet reached.

The focus of the argumentation view is twofold: representation of legal arguments but also the structure of each document in its elements and logical structure: factual information, evidence, arguments, conclusions, etc.[29] An emphasis has been placed on the formalization of argumentation in the 1990s, after the modest success of legal logic. Following the legal theory work on theories on legal argumentation (e.g.[30]), formalisation of legal arguments stays in the focus since the 1990s. The elements of argumentation are systematized in a logical argumentation structure. Arguments are an essential tool of legal work, because thesis, antithesis and synthesis are relevant in each case due to the dialectical roles of plaintiff, defendant and judge.

4. Methodology

For producing this enormous structural analysis, a lot of resources are needed in case of manual work, e.g. for searching, reading, interpreting and understanding of the legal text corpus. The standard methodology is to locate, read, interpret and understand the «legal stuff», taking into account the legal interpretation and reasoning methods in a dynamic world of concepts. Other elements are also considered, e.g. social context, legal authorities, sophisticated methods of interpretation etc. At present, financial constraints restrict deep analysis to areas of legal disputes or theoretical interest. Stronger use of AI & law can be the solution to this knowledge acquisition problem.

The most significant add-on of legal informatics is the revolution in legal search by the use of search engines. Legal information providers follow law-making, rendering of new judgements or decisions, publishing of new literature etc. They take care of all relevant changes of the particular legal system, index these materials in full-text and, if possible, add metadata and provide powerful search engines. The resulting – in the best case – 4 views are very helpful for the lawyer but do not substitute a full legal analysis, e.g. a conceptual and logical analysis using all interpretation methods. Here, the fundamental statement of SOWA applies: the terminology is to be developed and to be brought into a strong logical structure.

---

[26] Wikipedia EN, Cyc, https://en.wikipedia.org/wiki/Cyc (last accessed: 12 January 2016).

[27] T. BERNERS-LEE ET AL., The Semantic Web, *Scientific American* 284 (5), pages 34–53, 2001.

[28] C. BRUNSCHWIG, Multisensory Law and Legal Informatics – A Comparison of How these Legal Disciplines Relate to Visual Law. In: A. Geist, C. R. Brunschwig, F. Lachmayer, G. Schefbeck (eds.), *Strukturierung der Juristischen Semanik – Structuring Legal Semantics, Festschrift für Erich Schweighofer*, pages 573–668, Editions Weblaw, Bern 2011.

[29] K. D. ASHLEY, Modeling Legal Argument. Reasoning with Cases and Hypotheticals, MIT Press, Cambridge 1990.

[30] R. ALEXY, Theory der juristischen Argumentation (Theory of Legal Argumentation), Frankfurt am Main 1983.

AI & law methodology allows a representation according to the different needs of the users: Handbook, Dynamic Electronic Legal Commentary (DynELC), citizen information or case-related synthesis. AI & law methods can produce information elements for a comprehensive picture. Starting from the text corpus in XML format, legal data science methods drawn from many fields within the broad areas of mathematics, statistics, and information technology: pattern recognition and learning, machine learning, probability models, statistical learning, visualization, data warehousing, etc. AI & law has developed and refined these methods for legal output. It has to be noted that the required accuracy for practical use has very often not been achieved yet.

4.1. Information elements

Document categorisation: An assignment of one of the more than 200 document types is relevant for the role of the document in the legal systems and determines some of its attributes (e.g. preparatory act, law, implementing act etc.). This categorisation can be further developed, e.g. in doing also privilege classification[31].

Document segmentation: A very important but often neglected part is the segmentation of a legal document into its structural elements, e.g. author, title, material rules, procedural rules, facts, arguments, rulings etc. The semantic meaning is much stronger if the document part is also taken into account.

Citations: The automatic generation of hypertext links describes the network between documents but also its relevance. Here, the strongest use of AI & law can be noticed in practice (e.g. the Australasian Legal Information Institute AustLII uses hyperlinks as a main method of analysis support besides full text retrieval[32].

Temporal relations: The automatic generation of temporal relationships analyses the temporal dimensions: existence, force, efficacy and applicability. It is standard in information retrieval but not sufficiently implemented in the AI & law environment. A concept exists with the work of SCHARF[33].

Document description and summary: A lexical ontology, e.g. a thesaurus, remains critical for a condensed short content description. Multilingual versions like Eurovoc or the Swiss legal thesaurus are very helpful for first handling of documents in a foreign language. Much work has been done already in the 1990es (e.g. FLEXICON and KONTERM, see for an analysis[34]).

Logical and argument analysis: Given the document segmentation, textual rules and arguments can be also formally represented in a pre-defined model, e.g. in very useful practical flow charts (activity diagrams) or argumentation structures. Here, much work has still to be done.

Visualisations: Recently, traditional pictorial representations are supplemented with model-based visualizations. Visualisations are very helpful for representation of «soft» relations between documents, concepts and arguments[35].

User assessment: Modern search engines can (semi)automatically generate user's perspectives.[36]

---

[31] J. K. VINJUMUR, Evaluating Expertise and Sample Bias Effects for Privilege Classification in E-Discovery. *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law ICAIL 2015*, pages 119–128, ACM, New York 2015.

[32] Australasian Legal Information Institute AustLII: http://www.austlii.edu.au/ (lastly accessed: 12 January 2016).

[33] See footnote 18.

[34] E. SCHWEIGHOFER, Legal Knowledge Representation. Kluwer Law International 1999.

[35] P. WAHLGREN, Legal Risk Analysis. A Proactive Legal Method, Stockholm 2013.

[36] Q. LU, J. G. CONRAD, M. DAHN, W. M. KEENAN, Systems, methods, and interfaces for extending legal search results. US States Patent No. 9,177,050 USPTO Patent Full-Text and Image Database: http://patft.uspto.gov (lastly accessed: 5 February 2016), 2015.

4.2. AI & law methods

Due to space reasons, only a short overview on the state of the methodology can be given. The focus is on pattern recognition, thesaurus generation, machine learning, natural language processing, ranking, legal ontologies, legal logic, legal argumentation and visualization.

The standard representation of documents in XML format has to be improved by use of a standard part-of-speech tagger and a parser. Thus, pattern matching, machine learning and natural language processing[37] can be strongly improved for document categorisation, document segmentation, citation analysis[38], temporal relations[39] and thesaurus generation.

It is a hermeneutic process of knowledge generation and re-use, using a knowledge base and the text corpus. Patterns of document types, document segments, citation structures, temporal relations and concepts are matched with the text corpus resulting in a raw analysis of a document. The knowledge base has to be refined and improved using machine learning techniques. A learning circle exists once such a system is implemented for daily use.

The output is presented in a pre-defined formal model (e.g. types of citations with relevance ranking). Further, this semi-automatic approach delivers also relevant feedback from authors and users for improving the knowledge base[40]. The research on the quality of legislation is also relevant[41]. A strong support can be expected from the study of natural language features[42].

Model-based visualizations are based on grammars, i.e. strict specifications for composing visualizations based on sets of pre-defined elements and relation (e.g. the PICTMOD of Fill[43]).

Ranking is used as a mechanism for establishing a relationship of relevance for a particular information or analysis. In law, it is used in legal information retrieval, however, with still insufficient success compared to the Google algorithm (cf.[44]). Ranking gives some guidance concerning the relations of a document vs. requirements, document in the corpus, document in the citations network, document in the timeline etc.

The logical analysis and the argumentation analysis require a rule model and an argumentation model. Some support can be achieved if text is transferred into logical rules (e.g. the well-known example of Softlaw/Oracle[45] or recently[46]). Flow-charts (activity diagrams) are also very helpful in practice. Recent work on argumenta-

---

[37] See foonote 34.

[38] E. Schweighofer, D. Scheithauer, The Automatic Generation of Hypertext Links in Legal Documents. In *Proceedings of Database and Expert Systems Applications, 7th International Conference, DEXA'96*, Zurich 1996, Lecture Notes in Computer Science 1134, pages 889–898, Springer, Berlin 1996.

[39] See footnote 18.

[40] See footnotes 4 and 34.

[41] M. Curtotti et al., Machine Learning for Readability of Legislative Sentences, In: *Proceedings of of the Fifteenth International Conference on Artificial Intelligence and Law, ICAIL2015,* pages 53–62, ACM, New York 2015.

[42] Bernhard Waltl, Florian Matthes, Supporting the Legal Subsumption Process: Determination of Concreteness and Abstractness in German Laws using Lexical Knowledge, In: *Proceedings of MWAIL2015*, pages 75–88, OCG, Vienna 2015.

[43] H.-G. Fill, Bridging Pictorial and Model-based Creation of Legal Visualizations: The PICTMOD Method. In: E. Schweighofer, F. Kummer, W. Hötzendorfer (eds.), *Proceedings of International Legal Informatics Conference IRIS2015*, pages 443–447, OCG, Vienna 2015.

[44] A. Geist, Relevanzsortierung bei Rechtsdatenbanken (relevance ranking in legal databases), Ph.D thesis (to be submitted), University of Vienna 2016.

[45] S. Dayal, M. Harmer, P. Johnson, D. Mead, Beyond Knowledge Representation: Commercial Uses for Legal Knowledge Bases. In: *Proceedings of the Fourth International Conference on Artificial Intelligence and Law ICAIL1993*, pages 167–174, ACM, New York 2015.

[46] M.D. Islam, G. Governatori, RuleOMS: A Rule-Based Online Management System. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law ICAIL 2015*, pages 187–191, ACM, New York 2015.

tion analysis can be found in the ICAIL2015 proceedings[47]. The logic programming is powerful enough to be a decisive support tool for the implementation of laws. Well-structured flowcharts help in decision-making, esp. when thousands of legal rules must be processed. However, two important elements should be added: a model of the layers of time[48]; further, a differentiation according to distinct applicability of norms to legal subjects.

At the moment, a transitional period can be noticed characterised by semi-automation. Computers can be very supportive but in the very end a human being has to check the proper context and quality of the work. A fruitful cooperation between men and machines has to be developed in order to deploy and improve the methodology of legal data science.

So far, all these tools of data analysis are insufficiently developed and used in the legal domain. It is sad that legal authors rarely use such tools in order to generate synthesis works like commentaries. In our view, the AI & law methodology is sufficiently strong that legal publishers should move on and provide such a toolbox for its authors.

5. Conclusions

Legal data science looks back to more than 50 years of experience in the analysis of legal text corpora (albeit under different names, in particular legal information retrieval or AI & law). Further, the long tradition of legal theory research provides some guidance. Present analysis consists in creating 8 views using AI & law methodology for producing automatically or semi-automatically related input for further intellectual work. The stronger use of computational methods speeds up legal analysis and allows a faster and more efficient assessment of the legal materials. In this transitional period, a fruitful cooperation between men and machines has to be developed in order to deploy and improve the methodology of legal data science. A huge potential remains unused for the more efficient analysis of legal text corpora.

This paper was also presented at the JURIX2015 CEILI Workshop on Legal Data Analysis in December 2015.

---

[47] E.g. J. Park, C. Blake, C. Cardie,. Toward Machine-assisted Participation in eRulemaking: An Argumentation Model of Evaluability. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law, ICAIL2015*, pages 206–210, ACM, New York, 2015; M. Dragoni, G. Governatori, S. Villata, Towards Automated Rules Generation from Natural Language Legal Texts. In: *ICAIL 2015 Workshop on Automated Detection, Extraction and Analysis of Semantic Information in Legal Texts*, 2015; G. Governatori, Thou Shalt is not You Will. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law ICAIL 2015*, pages 63–68, ACM, New York 2015; L. Al-Abdulkarin, K. Atkinson, T. Bench-Capon, Factors, Issues and Values: Revisiting Reasoning with Cases. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law ICAIL 2015*, pages 3–12, ACM, New York 2015.

[48] J. Scharf, rOWLer – A hybrid rule engine for legal reasoning. In: E. Schweighofer, F. Kummer, W. Hötzendorfer (eds.), *Proceedings of International Legal Informatics Conference IRIS2015*, pages 155–162, OCG, Vienna 2015.