

LEGAL INFORMATICS

This groundbreaking work offers a first-of-its-kind overview of legal informatics, the academic discipline underlying the technological transformation and economics of the legal industry. Edited by Daniel Martin Katz, Ron Dolin, and Michael J. Bommarito, and featuring contributions from more than two dozen academic and industry experts, chapters cover the history and principles of legal informatics and background technical concepts – including natural language processing and distributed ledger technology. The volume also presents real-world case studies that offer important insights into document review, due diligence, compliance, case prediction, billing, negotiation and settlement, contracting, patent management, legal research, and online dispute resolution. Written for both technical and non-technical readers, *Legal Informatics* is the ideal resource for anyone interested in identifying, understanding, and executing opportunities in this exciting field.

Daniel Martin Katz is Professor of Law, Illinois Tech – Chicago Kent College of Law where he directs The Law Lab. He also serves as an external affiliated faculty at CodeX, the Stanford Center for Legal Informatics and the Academic Director of the Bucerius Center for Legal Technology and Data Science. A scientist and technologist, Professor Katz applies an innovative, polytechnic approach to teaching law to help create lawyers for today's biggest societal challenges. Both his teaching and scholarship integrate science, technology, engineering, and mathematics.

Ron Dolin is a Senior Research Fellow and Lecturer on Law at Harvard Law School, focusing on the impact of technology on the practice and nature of law. He received a B.A. in math and physics, a Ph.D. in computer science, and a J.D. He has worked at JPL, CERN, and Google, and is a licensed attorney in California. Dolin's research includes developing legal quality metrics, examining the impact of standardized benchmarks on the legal system, and analyzing the legal market from the perspective of The Innovator's Dilemma.

Michael J. Bommarito is an entrepreneur, educator, and investor in the legal and technology industries. His experience spans R&D, technology, business, and operations – ranging from top Am Law firms and \$B+ AUM investment firms to idea-stage startups. He is affiliated with the University of Michigan, Stanford University, Michigan State College of Law, and the Illinois Tech – Chicago Kent College of Law. His research has been published in *Science*, *Physica A*, *Artificial Intelligence and Law*, and *Quantitative Finance*.

Cambridge University Press
978-1-107-14272-5 — Legal Informatics
Edited by Daniel Martin Katz , Ron Dolin , Michael J. Bommarito
Frontmatter
[More Information](#)

Legal Informatics

Edited by

DANIEL MARTIN KATZ

Illinois Tech – Chicago Kent Law
Bucerius Law School
Stanford CodeX

RON DOLIN

Harvard Law School

MICHAEL J. BOMMARITO

Stanford CodeX



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
978-1-107-14272-5 — Legal Informatics
Edited by Daniel Martin Katz , Ron Dolin , Michael J. Bommarito
Frontmatter
[More Information](#)

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107142725

DOI: 10.1017/9781316529683

© Cambridge University Press 2021

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2021

A catalogue record for this publication is available from the British Library.

ISBN 978-1-107-14272-5 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third-party internet websites referred to in this publication
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

Contents

<i>List of Figures</i>	<i>page</i> viii
<i>List of Tables</i>	xi
<i>List of Contributors</i>	xii
 PART I INTRODUCTION TO LEGAL INFORMATICS	 1
1.1 Motivation and Rationale for this Book	3
Michael J. Bommarito II, Daniel Martin Katz, and Ron Dolin	
1.2 Technology Issues in Legal Philosophy	5
Ron Dolin	
1.3 The Origins and History of Legal Informatics	24
Michael J. Bommarito II	
 PART II LEGAL INFORMATICS: BUILDING BLOCKS AND CORE CONCEPTS	 31
 A. INFORMATION REPRESENTATION, PREPROCESSING, AND DOCUMENT ASSEMBLY	 33
2.1 Representation of Legal Information	35
Katie Atkinson	
2.2 Information Intermediation	41
Ron Dolin	
2.3 Preprocessing Data	55
Michael J. Bommarito II	
2.4 XML in Law: The Role of Standards in Legal Informatics	61
Ron Dolin	
2.5 Document Automation	69
Marc Lauritsen	

B. ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, NATURAL LANGUAGE PROCESSING, AND BLOCKCHAIN	85
2.6 AI + Law: An Overview Daniel Martin Katz	87
2.7 Machine Learning and Law Daniel Martin Katz and John J. Nay	94
2.8 Natural Language Processing for Legal Texts John J. Nay	99
2.9 Introduction to Blockchain and Cryptography Nelson M. Rosario	114
C. PROCESS IMPROVEMENT, GAMIFICATION, AND DESIGN THINKING	121
2.10 Legal Informatics-Based Technology in Broader Workflows Kenneth A. Grady	123
2.11 Gamification of Work and Feedback Systems Stephanie Kimbro	136
2.12 Introduction to Design Thinking for Law Margaret Hagan	155
D. EVALUATION	177
2.13 Measuring Legal Quality Ron Dolin	179
PART III USE CASES IN LEGAL INFORMATICS	201
A. CONTRACTS AND PATENTS	203
3.1 Contract Analytics Noah Waisberg	205
3.2 Contracts as Interfaces: Visual Representation Patterns in Contract Design Helena Haapio and Stefania Passera	213
3.3 Distributed Ledgers, Cryptography, and Smart Contracts Nina Gunther Kilbride	239
3.4 Patent Analytics: Information from Innovation Andrew W. Torrance and Jevin D. West	257
B. LITIGATION AND E-DISCOVERY	289
3.5 The Core Concepts of E-discovery Jonathan Kerry-Tyerman and A. J. Shankar	291

3.6	Predictive Coding in E-discovery and the NexLP Story Engine Irina Matveeva	315
3.7	Examining Public Court Data to Understand and Predict Bankruptcy Case Results Warren E. Agin	335
	C. LEGAL RESEARCH, GOVERNMENT DATA, AND ACCESS TO LEGAL INFORMATION	355
3.8	Fastcase, and the Visual Understanding of Judicial Precedents Ed Walters and Jeff Asjes	357
3.9	Mining Information from Statutory Texts in a Public Health Domain Kevin D. Ashley	371
3.10	Gov2Vec: A Case Study in Text Model Application to Government Data John J. Nay	393
3.11	Representation and Automation of Legal Information Katie Atkinson	397
	D. DISPUTE RESOLUTION AND ACCESS TO JUSTICE	407
3.12	Online Dispute Resolution Dave Orr and Colin Rule	409
3.13	Access to Justice and Technology: Reaching a Greater Future for Legal Aid Ronald W. Staudt and Alexander F. A. Rabanal	416
3.14	Designing Legal Experiences: Online Communication and Resolution in Courts Maximilian A. Bulinski and J. J. Prescott	430
	PART IV LEGAL INFORMATICS IN THE INDUSTRIAL CONTEXT	449
	A. CHALLENGES FACING INNOVATION IN LAW	451
4.1	Adaptive Innovation: The Innovator’s Dilemma in Big Law Ron Dolin and Thomas Buley	453
4.2	Legal Data Access Christine Bannan	467
	B. LARGE FIRM AND CORPORATE LEGAL INFORMATICS CASE STUDIES	481
4.3	A History of Knowledge Management at Littler Mendelson Scott Rechtschaffen	483
4.4	Legal Operations at Google Mary O’Carroll and Stephanie Kimbro	501

Figures

1.3.1	This nilometer on Elephantine Island in Aswan, Egypt is one of the first applications of informatics in the law	<i>page</i> 25
1.3.2	<i>Corpus Juris Civilis</i> , republished in 1627 with commentary by Johann Fehe	27
2.1.1	CATO abstract factor hierarchy	39
2.2.1	Term vector spaces: (a) simple query; (b) complex query	49
2.3.1	A simplified, illustrative workflow for preprocessing data	56
2.5.1	Functions of document automation	70
2.5.2	Realms of knowledge management	70
2.6.1	Subfields within artificial intelligence	89
2.8.1	Example of a network of similar bills under consideration by Congress	101
2.8.2	Number of presidential proclamations or determinations, trivial proclamations, memoranda, and executive orders, from 1928 through 2015	106
2.8.3	Positive correlations between topics represented as lines	107
2.8.4	Part-of-speech tagging for a sentence in a state bill	108
2.8.5	Named entity recognition for the last part of the bill sentence	109
2.8.6	Syntactic dependencies	110
2.8.7	Co-reference resolution	110
2.8.8	Word2Vec algorithm	112
2.10.1	Anna’s process	125
2.10.2	Chart of email steps	126
2.10.3	Anna’s email process	126
2.10.4	Example workflow	129
2.10.5	Process stick for Anna’s activities	132
2.12.1	Six levels of legal design work that can improve users’ experience with the legal system, from least ambitious to most	173
2.13.1	88.4% of what? Failed to settle on 11.6% of losing cases?	180
2.13.2	Additive scoring (a) vs. (unweighted) multiplicative scoring (b)	193
2.13.3	Weighted multiplicative scoring	195
2.13.4	Geographical relevance of 0, 0.05, and 1	196
3.2.1	The emerging view of the purpose and functions of contracts	217
3.2.2	Linear timelines	226
3.2.3	Circular timeline	227
3.2.4	Processes with different paths	228
3.2.5	Variables over time	228
3.2.6	Synchronous progress of multiple processes or events	229
3.2.7	Flowchart	230
3.2.8	Table with bulleted lists	231

3.2.9	Table with color-coding and icons to show key differences	232
3.2.10	Swimlane	233
3.2.11	Icons in a tenancy agreement	235
3.2.12	Delivery diagram	236
3.2.13	Two different timelines, two different understandings	237
3.2.14	Decision tree evaluating average expected value of alternative decisions	238
3.3.1	A cryptographic hash	240
3.3.2	Linked timestamping storage in hash trees (Merkle trees)	242
3.3.3	Linked timestamping of case law	242
3.4.1	US Patent No. 6,164,870	259
3.4.2	A CIPO patent landscape of shale oil and gas technologies	274
3.4.3	Federal judicial districts by mean importance of patents litigated there	276
3.4.4	An example citation network	285
3.4.5	Automatic categorization	286
3.5.1	The EDRM	292
3.6.1	Visualization of the machine learning process	321
3.6.2	Mathematical objects in SVM with decision boundaries	323
3.6.3	A news article containing last names, nicknames, abbreviations, and metonymy	324
3.6.4	LSI-created word associations	326
3.6.5	Subdivisions within a cluster	328
3.6.6	Nested cluster diagram	329
3.6.7	Dendrogram	329
3.6.8	Entities extracted from an article about airline fare increases	330
3.6.9	NexLP Enron email communication network and content summary for emails discussing Enron	332
3.6.10	Example of emails under investigation	333
3.6.11	Failure of simple keyword search for “William Brown”	333
3.6.12	Story Engine summarizes data and discovers relevant details	333
3.7.1	This chart shows the number of cases in the data set for each disposition code	341
3.7.2	Case volume and success rate by judicial district, grouped vertically by federal circuit	343
3.7.3	Example of a decision tree	344
3.7.4	A small section of the final decision tree	345
3.7.5	Importance scores for model features	346
3.7.6	Confusion matrix for the three test sets, combined	347
3.7.7	Probability output to percentage conversion graph	349
3.8.1	A problem search	360
3.8.2	Fastcase search filters	361
3.8.3	Boolean search	361
3.8.4	Image from Fastcase’s Interactive Timeline	367
3.8.5	Y-axis changed to show court level	368
3.8.6	“Same-sex marriage” Interactive Timeline, with arrowing indicating <i>Lawrence v. Texas</i>	369
3.8.7	Close-up of Figure 3.8.6, “Same-sex marriage” case relevance	370
3.9.1	LENA statutory network comparing Texas and Pennsylvania statutes on epidemic emergencies with infectious diseases	377
3.9.2	Heat map of agent strength in statutory networks	378
3.9.3	Schematic view of phases of coding text by the SPH team	379
3.9.4	Sample public health emergency statutory provision	380
3.9.5	Results of initial machine learning	383

3.9.6	Results of successively applying more states’ classification models	387
3.9.7	Sample comparison of statutory provision texts	390
3.9.8	Interactive ML tool screen	390
3.9.9	Results of evaluation of the interactive ML approach	391
3.10.1	Sentence probabilities across bills for first data (a) and last data (b)	396
3.11.1	CATO abstract factor hierarchy	398
3.11.2	IBP logical model	400
3.11.3	Example abstract argumentation framework	401
3.11.4	Visualization of automobile exception domain addressing the question “Can an automobile be searched without a warrant?”	406
3.13.1	Conditional logic to determine judicial district in IL	421
3.13.2	Conditional logic to determine the end-user’s full name	422
3.13.3	Example of map feature (i.e., decision tree)	422
3.13.4	Report feature – readability score	423
3.13.5	A2J Guided Interview progress bar in Spanish	423
3.13.6	Example of a question screen in A2J Author	423
3.13.7	A2J Author usage on LHI, 2008–2016	427
3.14.1	Matterhorn litigant access	434
3.14.2	Eligibility for review	435
3.14.3	Judge’s reviews of a litigant’s request	436

Tables

2.5.1	Document automation range of functions	<i>page</i> 71
2.5.2	Document assembly products, past and present	77
2.5.3	Worksheet for document automation goals	82
2.10.1	Process study sheet	125
3.2.1	Examples of pattern structures	222
3.2.2	Categorization of contract visualization patterns	223
3.5.1	Hash values for the Gettysburg Address	303
3.6.1	LSI-computed areas of interest in Enron emails	327
3.6.2	Main named entity types	329
3.7.1	Prediction accuracy of model for different groupings of probability numbers	348
3.9.1	Comparing similarly purposed provisions from Pennsylvania (left) and Florida (right)	385
3.9.2	Results of models trained on Florida (FL), Pennsylvania (PA), and combined (FL + PA) data sets applied to PA (left) and FL (right) test sets	386
3.10.1	Synthetic summary for three topics, “Enacted” or “Failed,” in the House or the Senate	395
3.11.1	Base-level factors in CATO	399
3.11.2	CATO as ADF	402
3.11.3	IBP logical model as an ADF	402

Contributors

Warren E. Agin Adjunct Professor, Boston College Law School; Managing Director for Digital Strategy and Solutions, Elevate Services

Kevin D. Ashley Professor of Law, University of Pittsburgh

Jeff Asjes Product Manager, Fastcase

Katie Atkinson Professor of Computer Science, University of Liverpool

Christine Bannan Policy Counsel, Open Technology Institute, New America

Michael J. Bommarito II Fellow, CodeX – Stanford Center for Legal Informatics

Thomas Buley Vice-President of Product, Guideline

Maximilian A. Bulinski Research Fellow, University of Michigan Law School

Ron Dolin Lecturer on Law, Harvard Law School; Senior Research Fellow, Center on the Legal Profession, Harvard Law School

Kenneth A. Grady Adjunct Professor, Michigan State University College of Law

Helena Haapio Associate Professor of Business Law, University of Vaasa; International Contract Counsel, Lexpert Ltd.

Margaret Hagan Director Legal Design Lab, Stanford Law School Lecturer, Stanford Institute of Design (d.school)

Daniel Martin Katz Professor of Law, Illinois Tech – Chicago Kent College of Law Academic Director, Bucerius Center for Legal Technology & Data Science; Affiliated Faculty, CodeX – Stanford Center for Legal Informatics

Jonathan Kerry-Tyerman Vice-President of Business Development, Everlaw

Nina Gunther Kilbride Principal, Corpening Labs

Stephanie Kimbro Former Research Fellow, Stanford Law School; Independent Consultant, Virtual Law Practice

Marc Lauritsen President, Capstone Practice Systems

Irina Matveeva Adjunct Professor, Illinois Tech; Head of Machine Learning, NexLP

- John J. Nay** CEO & Co-Founder, Skopos Labs
- Mary O’Carroll** Director of Legal Operations, Google; President, CLOC
- Dave Orr** Senior Product Manager, Google
- Stefania Passera** Legal Tech Lab, Helsinki University
- J. J. Prescott** Professor of Law, University of Michigan Law School
- Alexander F. A. Rabanal** Associate Director, The Law Lab @ Illinois Tech – Chicago Kent College of Law
- Scott Rechtshaffen** Chief Knowledge Officer, Littler Mendelson
- Nelson M. Rosario** Adjunct Professor, Illinois Tech – Chicago-Kent College of Law
Principal, Smolinski Rosario Law P.C.
- Colin Rule** Vice-President for Online Dispute Resolution, Tyler Technologies
- A. J. Shankar** Founder and CEO, Everlaw
- Ronald W. Staudt** Professor of Law Emeritus, Illinois Tech – Chicago Kent College of Law
- Andrew W. Torrance** Paul E. Wilson Distinguished Professor of Law, University of Kansas School of Law
- Noah Waisberg** CEO and Co-founder, Kira Systems
- Ed Walters** CEO and Co-founder, Fastcase
- Jevin D. West** Associate Professor, Information School, University of Washington

Cambridge University Press
978-1-107-14272-5 — Legal Informatics
Edited by Daniel Martin Katz , Ron Dolin , Michael J. Bommarito
Frontmatter
[More Information](#)
