

## ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS

### *New Tools for Law Practice in the Digital Age*

The field of artificial intelligence (AI) and the law is on the cusp of a revolution that began with text analytic programs like IBM's Watson and Debater and the open-source information management architectures on which they are based. Today, new legal applications are beginning to appear, and this book – designed to explain computational processes to non-programmers – describes how they will change the practice of law, specifically by connecting computational models of legal reasoning directly with legal text, generating arguments for and against particular outcomes, predicting outcomes, and explaining these predictions with reasons that legal professionals will be able to evaluate for themselves. These legal apps will support conceptual legal information retrieval and enable cognitive computing, enabling a collaboration between humans and computers in which each performs the kinds of intelligent activities that they can do best. Anyone interested in how AI is changing the practice of law should read this illuminating work.

**Dr. Kevin D. Ashley** is a Professor of Law and Intelligent Systems at the University of Pittsburgh, Senior Scientist, Learning Research and Development Center, and Adjunct Professor of Computer Science. He received a B.A. from Princeton University, a JD from Harvard Law School, and Ph.D. in computer science from the University of Massachusetts. A visiting scientist at the IBM Thomas J. Watson Research Center, NSF Presidential Young Investigator, and Fellow of the American Association for Artificial Intelligence, he is co-Editor-in-Chief of *Artificial Intelligence and Law* and teaches in the University of Bologna Erasmus Mundus doctoral program in Law, Science, and Technology.

# Artificial Intelligence and Legal Analytics

NEW TOOLS FOR LAW PRACTICE IN  
THE DIGITAL AGE

**KEVIN D. ASHLEY**

University of Pittsburgh School of Law



**CAMBRIDGE**  
UNIVERSITY PRESS

Cambridge University Press  
978-1-107-17150-3 — Artificial Intelligence and Legal Analytics  
Kevin D. Ashley  
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](http://www.cambridge.org)  
Information on this title: [www.cambridge.org/9781107171503](http://www.cambridge.org/9781107171503)  
DOI: 10.1017/9781316761380

© Kevin D. Ashley 2017

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 2017  
Reprinted 2018

Printed in the United Kingdom by Clays, St Ives plc

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

*Library of Congress Cataloging-in-Publication Data*

ISBN 978-1-107-17150-3 Hardback  
ISBN 978-1-316-62281-0 Paperback

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.

Cambridge University Press  
978-1-107-17150-3 – Artificial Intelligence and Legal Analytics  
Kevin D. Ashley  
Frontmatter  
[More Information](#)

---

*For Alida, forever*

## Contents

<i>List of illustrations</i>	<i>page</i> xv
<i>List of tables</i>	xxi
<i>Acknowledgments</i>	xxiii
<b>PART I COMPUTATIONAL MODELS OF LEGAL REASONING</b>	<b>1</b>
<b>1 Introducing AI &amp; Law and Its Role in Future Legal Practice</b>	<b>3</b>
1.1. Introduction	3
1.2. AI & Law and the Promise of Text Analytics	4
1.3. New Paradigms for Intelligent Technology in Legal Practice	6
1.3.1. Former Paradigm: Legal Expert Systems	8
1.3.2. Alternative Paradigms: Argument Retrieval and Cognitive Computing	11
1.3.3. Toward the New Legal Apps	14
1.4. What Watson Can and Cannot Do	14
1.4.1. IBM's Watson	15
1.4.2. Question Answering vs. Reasoning	18
1.4.3. IBM's Debater Program	23
1.4.4. Text Analytic Tools for Legal Question Answering	27
1.4.5. Sources for Text Analytic Tools	30
1.5. A Guide to This Book	31
1.5.1. Part I: Computational Models of Legal Reasoning	32
1.5.2. Part II: Legal Text Analytics	33
1.5.3. Part III: Connecting Computational Reasoning Models and Legal Texts	34
1.6. Implications of Text Analytics for Students	35

<b>2</b>	<b>Modeling Statutory Reasoning</b>	38
2.1.	Introduction	38
2.2.	Complexities of Modeling Statutory Reasoning	39
2.2.1.	Semantic Ambiguity and Vagueness	39
2.2.2.	Syntactic Ambiguity	40
2.3.	Applying Statutory Legal Rules Deductively	42
2.3.1.	Running a Normalized Version on a Computer	43
2.3.2.	Predicate Logic	44
2.3.3.	Syntactic Ambiguity as Design Constraint	45
2.3.4.	The BNA Program	47
2.3.5.	Some Problems of Translating Statutes into Programs	48
2.4.	The Complexity of Statutory Interpretation and the Need for Arguments	52
2.4.1.	A Stepwise Process of Statutory Interpretation	53
2.4.2.	Other Sources of Legal Indeterminacy	54
2.5.	Management Systems for Business Rules and Processes	56
2.5.1.	Business Process Expert Systems	56
2.5.2.	Automating Business Process Compliance	60
2.5.3.	Requirements for a Process Compliance Language	62
2.5.4.	Connecting Legal Rules and Business Processes	65
2.5.5.	Example of Business Process Compliance Modeling	68
2.6.	Representing Statutory Networks	70
<b>3</b>	<b>Modeling Case-based Legal Reasoning</b>	73
3.1.	Introduction	73
3.2.	Relationship of Legal Concepts and Cases	74
3.2.1.	The Legal Process	74
3.2.2.	The Legal Process Illustrated	75
3.2.3.	Role of Legal Concepts	76
3.3.	Three Computational Models of Legal Concepts and Cases	77
3.3.1.	Prototypes and Deformations	78
3.3.2.	Dimensions and Legal Factors	81
3.3.3.	Exemplar-based Explanations	93
3.4.	Teleological Models of Case-based Legal Reasoning	97
3.5.	An Approach to Modeling Teleological Reasoning	100
3.5.1.	Teleology in Theory Construction	101
3.6.	Design Constraints for Cognitive Computing with Case-based Models of Legal Reasoning	104
<b>4</b>	<b>Models for Predicting Legal Outcomes</b>	107
4.1.	Introduction	107
4.2.	A Nearest Neighbor Approach to Automated Legal Prediction	108

4.3.	Introduction to Supervised Machine Learning	109
4.3.1.	Machine Learning Algorithms: Decision Trees	110
4.4.	Predicting Supreme Court Outcomes	111
4.4.1.	Features for Predicting Supreme Court Outcomes	112
4.4.2.	Applying Supervised Machine Learning to SCOTUS Data	112
4.4.3.	Evaluating the Machine Learning Method	113
4.4.4.	Machine Learning Evaluation Measures and Results	114
4.5.	Predicting Outcomes with Case-based Arguments	114
4.5.1.	Prediction with CATO	115
4.5.2.	Issue-based Prediction	115
4.5.3.	IBP's Prediction Algorithm	117
4.5.4.	Evaluating IBP's Predictions	119
4.6.	Prediction with Underlying Values	121
4.7.	Prediction based on Litigation Participants and Behavior	123
4.8.	Prediction in Cognitive Computing	125
<b>5</b>	<b>Computational Models of Legal Argument</b>	<b>127</b>
5.1.	Introduction	127
5.1.1.	Advantages of CMLAs	128
5.2.	The Carneades Argument Model	129
5.3.	An Extended Example of a CMLA in Action	131
5.3.1.	Family Law Example with Carneades	132
5.3.2.	Arguing with Defeasible Legal Rules	134
5.3.3.	Integrating Arguing with Cases and Rules	135
5.4.	Computational Model of Abstract Argumentation	139
5.5.	How CMLAs Compute Winners and Losers	141
5.5.1.	Resolving Conflicting Arguments about Facts	142
5.5.2.	Resolving Conflicting Arguments about Values	143
5.5.3.	Resolving Conflicting Arguments about Legal Rules	144
5.6.	How Practical are Computational Models of Legal Argument?	144
5.6.1.	Role of Proof Standards in CMLAs	145
5.6.2.	Integrating Probabilistic Reasoning into CMLAs	147
5.7.	Value Judgment-based Argumentative Prediction Model	149
5.7.1.	VJAP Domain Model	150
5.7.2.	VJAP Values Underlying Trade Secret Regulation	151
5.7.3.	VJAP Argument Schemes	154
5.7.4.	VJAP's Argument-based Predictions	156
5.7.5.	VJAP Program Evaluation	158
5.8.	Computational Model of Evidentiary Legal Argument	160
5.9.	Computational Models of Legal Argument as a Bridge	164

<b>PART II</b>	<b>LEGAL TEXT ANALYTICS</b>	169
<b>6</b>	<b>Representing Legal Concepts in Ontologies and Type Systems</b>	171
6.1.	Introduction	171
6.2.	Ontology Basics	172
6.3.	Sample Legal Ontologies	174
6.3.1.	The e-Court Ontology	174
6.3.2.	van Kralingen's Frame-based Ontology	176
6.4.	Constructing Legal Ontologies	178
6.5.	Ontological Support for Statutory Reasoning	181
6.6.	Ontological Support for Legal Argumentation	185
6.6.1.	A Target Application for Legal Argument Ontology	185
6.6.2.	An Ontology for the Argument Microworld	190
6.6.3.	Limits for Automating Legal Argumentation through Ontological Support	198
6.6.4.	Ontological Support for Cognitive Computing in Legal Argumentation	201
6.7.	Type Systems for Text Analytics	202
6.7.1.	Defining a Type System	202
6.7.2.	Type System Example: DeepQA	203
6.8.	LUIMA: A Legal UIMA Type System	204
6.9.	LUIMA Annotations can Support Conceptual Legal Information Retrieval	208
<b>7</b>	<b>Making Legal Information Retrieval Smarter</b>	210
7.1.	Introduction	210
7.2.	Current Legal Information Retrieval Services	211
7.3.	An Example of Using Commercial Legal IR Systems	212
7.4.	How Legal IR Systems Work	214
7.5.	IR Relevance Measures	216
7.5.1.	Boolean Relevance Measure	216
7.5.2.	Vector Space Approach to Relevance	217
7.5.3.	Probabilistic Model of Relevance	218
7.6.	Assessing Legal IR Systems	221
7.7.	Recent Developments in Legal IR Systems	223
7.8.	Comparing Legal IR and CMLAs	226
7.9.	Improving Legal IR with AI & Law Approaches	226
7.9.1.	Integrating Legal Ontologies and IR	227
7.9.2.	Integrating Legal IR and AI & Law Relevance Measures	227



7.9.3. Augmenting Legal IR Relevance Assessment with Citation Networks	230
7.9.4. Detecting Concept Change	232
7.10. Conclusion	233
<b>8 Machine Learning with Legal Texts</b>	<b>234</b>
8.1. Introduction	234
8.2. Applying Machine Learning to Textual Data	234
8.3. A Basic Setup for Applying ML to Legal Texts	236
8.4. Machine Learning for e-Discovery	239
8.4.1. Litigators' Hypotheses in e-Discovery	240
8.4.2. Predictive Coding Process	241
8.4.3. Assessing Predictive Coding Effectiveness	243
8.4.4. Other Open Issues in Predictive Coding	246
8.4.5. Unsupervised Machine Learning from Text	247
8.5. Applying ML to Legal Case Texts in the History Project	248
8.5.1. History Project System Architecture	249
8.5.2. ML Algorithms: Support Vector Machines	251
8.5.3. History Project SVM	252
8.6. Machine Learning of Case Structures	253
8.7. Applying ML to Statutory Texts	254
8.7.1. Statutory Analysis	254
8.7.2. An Interactive ML Tool for Statutory Analysis	255
8.8. Toward Cognitive Computing Legal Apps	257
<b>9 Extracting Information from Statutory and Regulatory Texts</b>	<b>259</b>
9.1. Introduction	259
9.2. Research Overview Regarding Extracting Information from Statutory Texts	260
9.3. Automatically Extracting Functional Information from Statutory Provisions	262
9.3.1. Machine Learning to Extract Functional Types of Provisions	263
9.3.2. Text Classification Rules to Extract Functional Information	265
9.4. ML vs. KE for Statutory Information Extraction	266
9.5. Extracting Logical Rules from Statutes and Regulations	268
9.6. Extracting Requirements for Compliant Product Designs	270
9.6.1. Implementing Compliance with Extracted Regulations	272
9.6.2. Semiautomated Approaches to Improving Human Annotation for Compliance	272
9.7. Extracting Functional Information to Compare Regulations	275
9.7.1. Machine Learning for Constructing Statutory Networks	276
9.7.2. Applying an ML Algorithm for Statutory Texts	278

9.7.3.	Evaluating the ML Algorithm on Statutory Texts and Dealing with Sparse Training Data	280
9.7.4.	Applying LUIMA to Enrich Statutory Text Representation	282
9.8.	Conclusion	283
<b>10</b>	<b>Extracting Argument-Related Information from Legal Case Texts</b>	<b>285</b>
10.1.	Introduction	285
10.2.	Argument-Related Information in Legal Cases	286
10.3.	Extracting Legal Argument Claims	287
10.3.1.	Machine Learning to Classify Sentences as Propositions, Premises, and Conclusions	287
10.3.2.	Text Representation	288
10.3.3.	Applying Statistical Learning Algorithms	289
10.3.4.	Argument Grammar for Discourse Tree Structure	291
10.3.5.	Identifying Instances of Argument Schemes	293
10.4.	Extracting Argument-Related Legal Factors	294
10.4.1.	Three Representations for Learning from Text	294
10.4.2.	How Well Did SMILE Work?	297
10.4.3.	Annotating Factor Components	298
10.5.	Extracting Findings of Fact and Cited Legal Rules	299
10.5.1.	Applying the LUIMA Type System	299
10.5.2.	Preparing Gold Standard Cases	300
10.5.3.	LUIMA-Annotate	301
10.5.4.	Evaluating LUIMA-Annotate	304
10.6.	Annotation of Training Data	305
10.6.1.	Annotation in IBM Debater	306
10.6.2.	Annotation Protocols	308
10.6.3.	Computer-Supported Annotation Environments	308
<b>PART III CONNECTING COMPUTATIONAL REASONING MODELS AND LEGAL TEXTS</b>		<b>311</b>
<b>11</b>	<b>Conceptual Legal Information Retrieval for Cognitive Computing</b>	<b>313</b>
11.1.	Introduction	313
11.2.	State of the Art in Conceptual Legal IR	315
11.3.	LUIMA Architecture	316
11.3.1.	LUIMA-Search	316
11.3.2.	Reranking Documents with LUIMA-Rerank	320
11.4.	An Experiment to Evaluate LUIMA	321
11.4.1.	Evaluation Metrics	323
11.4.2.	LUIMA vs. CLIR	324

## Contents

xiii

11.5.	Continuing to Transform Legal IR into AR	327
11.5.1.	Connecting LARCCS and Legal IR Systems	328
11.5.2.	Querying for Cases with Extended Argument-Related Information	329
11.5.3.	New Legal Annotation Types	332
11.5.4.	Prospects for Annotating Expanded Legal Types	336
11.5.5.	Eliciting Users' Argument Needs	339
11.6.	Conceptual Information Retrieval from Statutes	342
11.6.1.	A Type System for Statutes	343
11.6.2.	Network Techniques for Conceptual Legal IR	345
11.6.3.	Conceptual Legal IR with Statutory Network Diagrams	346
11.7.	Conclusion	349
<b>12</b>	<b>Cognitive Computing Legal Apps</b>	<b>350</b>
12.1.	Introduction	350
12.2.	New Legal Apps on the Market	351
12.2.1.	Ross	351
12.2.2.	Lex Machina	353
12.2.3.	Ravel	353
12.3.	Bridging Legal Texts and Computational Models	354
12.4.	Cognitive Computing Apps for Testing Legal Hypotheses	354
12.4.1.	A Paradigm for CCLAs: Legal Hypothesis-Testing	355
12.4.2.	Targeted Legal Hypotheses	357
12.4.3.	Operationalizing Hypotheses	359
12.4.4.	Interpreting Hypotheses	361
12.5.	Challenges for Cognitive Computing Legal Apps	367
12.5.1.	Challenges: Automatically Annotating Legal Argument-Related Information	368
12.5.2.	Challenges: Manual Annotation of Training Instances	373
12.5.3.	Challenges: Query-Interface Design	379
12.6.	Detecting Opportunities for New Hypotheses and Arguments	381
12.7.	What to Do Next?	384
12.8.	Conclusion	390
	<i>Glossary</i>	393
	<i>Bibliography</i>	403
	<i>Index</i>	421

## Illustrations

1.1	Heuristic rules defining loss and strict liability (Waterman and Peterson, 1981)	9
1.2	ATN for offer and acceptance problems with four states: (o) no relevant legal relations; (1) offer pending; (2) contract exists; (12) contract exists and proposal to modify is pending (Gardner, 1987, p. 124)	19
1.3	Gardner’s heuristic method for distinguishing hard and easy legal questions (Gardner, 1987; Rissland, 1990)	22
1.4	Argument diagram of IBM Debater’s output for violent video games topic (root node) (see Dvorsky, 2014)	24
1.5	Diagram representing realistic legal argument involving violent video games topic	25
1.6	Architecture of text analyzer for legal documents including contracts. Dashed boxes show components for semantic analysis and conceptual information retrieval	29
2.1	Normalized versions of two alternative interpretations of the Louisiana statute and corresponding Prolog rules (bottom) (Allen and Engholm, 1978)	42
2.2	IRC section 354 and a normalized version (right) (see Allen and Engholm, 1978)	43
2.3	Flow chart for propositionalized IRC section 354 (see Allen and Engholm, 1978)	44
2.4	BNA provisions as represented in rules (Sergot et al., 1986)	47
2.5	BNA program output (excerpts) (Sergot et al., 1986, p. 376f)	49
2.6	Norm graphs for concluding “Legality” in section 4 (1) FDPA and “Effective Consent” (see Oberle et al., 2012, pp. 305–6, Figs. 13 and 14)	59

xvi	<i>Illustrations</i>	
2.7	Sample BPMN diagram of simple insurance claim process with business rule annotations (see Table 2.1) (Koetter et al., 2014, Fig. 2, p. 220)	60
2.8	Petri net representing simple producer–consumer resource allocation problem (see Kafura, 2011, p. 8)	66
2.9	Compliance system report of traces, rules, and tasks responsible for noncompliance (excerpts) (see Governatori and Shek, 2012)	69
2.10	Statutory network diagram comparing Pennsylvania (PA) and Florida (FL) statutory schemes re public health emergency surveillance: Circles indicate public health system actors and partners in FL and PA. Grey links indicate relationships present in both states; white links indicate legal relationships present in PA but not in FL (Sweeney et al., 2014)	71
3.1	Secrets-Disclosed-Outsiders dimension in Ashley (1990)	82
3.2	Hypo-style three-ply argument for the <i>Mason</i> case (see Ashley, 1990)	85
3.3	Hypo argument model with Venn diagram (Ashley, 1990)	86
3.4	Hypo claim lattice (Ashley, 1990)	87
3.5	Example of CABARET’s process for analyzing <i>Weissman v. IRS</i> , 751 F. 2d 512 (2d Cir. 1984) (Rissland and Skalak, 1991)	89
3.6	CATO Factor hierarchy (Aleven, 1997, 2003)	92
3.7	CATO argument downplaying/emphasizing distinction (Aleven, 2003)	92
3.8	GREBE semantic net representation of <i>Vaughn</i> case (Branting, 1991, 1999)	93
3.9	GREBE matches structure of <i>Vaughn</i> case to Jarek problem (Branting, 1991, 1999)	94
3.10	In-furtherance-of employment cases retrieved by GREBE for Jarek problem (Branting, 1991, 1999)	96
3.11	Excerpts of GREBE’s argument for Jarek problem (see Branting, 1991, 1999)	97
3.12	Which argument better accounts for teleological concerns? (Berman and Hafner, 1993)	99
3.13	Theory constructed from factor and value preferences (Bench-Capon and Sartor, 2003)	101
3.14	Argument as theory construction from factor and value preferences (see Bench-Capon and Sartor, 2003)	102
4.1	Projection of capital gains tax cases onto two dimensions (see Mackaay and Robillard, 1974)	109
4.2	Bail decisions data (a) from which decision tree is constructed (b)	110
4.3	Example features from Supreme Court database [S], the Segal–Cover Scores [SC], and feature engineering [FE] (Katz et al., 2014)	112
4.4	IBP domain model (Ashley and Brüninghaus, 2006)	116
4.5	IBP algorithm (Ashley and Brüninghaus, 2009)	117

*Illustrations*

xvii

4.6	Example of IBP’s prediction for MBL case	118
4.7	IBP vs. other prediction methods: results (Ashley and Brüninghaus, 2009)	119
4.8	Excerpts from a theory learned by AGATHA with <i>Mason</i> case as cfs (Chorley and Bench-Capon, 2005a)	121
5.1	A Carneades argument diagram (Gordon, 2008b,c; Gordon and Walton, 2009; Walton and Gordon, 2009; see Ashley, 2012)	131
5.2	Classical deduction (Gordon, 2008b,c; Gordon and Walton, 2009; Walton and Gordon, 2009)	133
5.3	Classical deduction cannot prove proposition and its opposite (see Gordon, 2008b,c; Gordon and Walton, 2009; Walton and Gordon, 2009)	133
5.4	Scheme for arguments from defeasible inference rules (see Gordon, 2008b,c; Gordon and Walton, 2009; Walton and Gordon, 2009)	134
5.5	Arguments (pro and con) with defeasible inference rules (Gordon, 2008b, c; Gordon and Walton, 2009; Walton and Gordon, 2009; see Ashley, 2012)	136
5.6	When the rules run out (see Gordon, 2008b, c; Gordon and Walton, 2009; Walton and Gardon, 2009)	136
5.7	Carneades case-based argument (con) (Gordon, 2008b, c; Gordon and Walton, 2009; Walton and Gordon, 2009; see Ashley, 2012)	138
5.8	Carneades case-based argument (pro) (Gordon, 2008b,c; Gordon and Walton, 2009; Walton and Gordon, 2009; see Ashley, 2012)	139
5.9	Attacking arguments as per VAF (see Atkinson and Bench-Capon, 2007, p. 113)	140
5.10	VJAP domain model (Grabmair, 2016)	150
5.11	Values protected by trade secret law: interests of plaintiffs in property and confidentiality (Grabmair, 2016)	152
5.12	Values protected by trade secret law: interests of general public in the usability of publicly available information and in fair competition (Grabmair, 2016)	153
5.13	VJAP Program output for <i>Dynamics</i> Case (excerpts) (Grabmair, 2016, pp. 59–60)	155
5.14	Statement and argument structure for reasoning about a Restatement issue with trade-offs in VJAP (Grabmair, 2016, p. 51)	157
5.15	DLF partial rule tree for vaccine decisions, showing three causation conditions of <i>Althen</i> (see Walker et al., 2011)	162
5.16	DLF extracted reasoning chains (excerpts) (see Walker et al., 2011)	163
6.1	Sample ontology for contract formation	172
6.2	Excerpts from e-Court ontology showing expansion of “reasoning object,” “agent,” and “juridical role.” Links are is-a unless otherwise noted (see Breuker and Hoekstra, 2004; Breuker et al., 2004; Van Engers et al., 2008)	175

6.3	Ontology in the DALOS system (excerpts) (see Francesconi et al., 2010)	179
6.4	Model of legal argument with hypotheticals (Ashley, 2009b)	188
6.5	Case frame for Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant)	191
6.6	Factor frame for Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant)	192
6.7	ILC frame for Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant)	196
6.8	Test frame for Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant)	197
6.9	Policy/Value frame for Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant)	197
6.10	Judge McCarthy's Test in <i>Popov v. Hayashi</i>	201
7.1	Three-dimensional vector space model	217
7.2	BN for the family-away problem (see Charniak, 1991, p. 52)	219
7.3	Inference network retrieval model (see Turtle, 1995, p. 33)	220
7.4	Architecture of SPIRE (Daniels and Rissland, 1997a)	228
7.5	Retrieval for "videocassette" (see Rose and Belew, 1991)	230
8.1	History Project system architecture (see Al-Kofahi et al., 2001)	249
8.2	SVM examples. Black dots are positive instances, gray dots are negative, and the white dot is "unknown" (see Noble, 2006, p. 1566)	252
8.3	Statutory analysis tool (Savelka et al., 2015)	256
9.1	Sample input/output of xmLegesExtractor (Francesconi, 2009, p. 66)	263
9.2	Multiclass SVM hyperplane example (see Francesconi and Passerini, 2007, p. 12)	264
9.3	Sample statutory sentences annotated for logical parts: antecedents $\langle A \rangle$ , consequents $\langle C \rangle$ , and topics $\langle T \rangle$ (Bach et al., 2013)	269
9.4	Term suggestions for annotation templates (see Yoshida et al., 2013)	274
9.5	Partitioning statutory provision into subtree chunks	278
9.6	Decision tree for classifying statutory text units as relevant ( <i>rel.</i> ) or irrelevant ( <i>irrel.</i> ) to the School of Public Health study	279
9.7	The box plot summarizes the results of all the experiments on each of the tasks for Florida (FL). Each box plot describes performance in terms of an F1-measure within a single experiment. The tasks included identifying: AA (acting agent), PR (prescription), AC (action), GL (goal), PP (purpose), ET (emergency type), RA (receiving agent), CN (condition), TF (time frame) (Savelka and Ashley, 2015)	282
10.1	Excerpt of argument tree structure extracted automatically from a case (see Moens et al., 2007, Fig. 5)	292
10.2	Overview of SMILE and IBP (see Ashley and Brünninghaus, 2009)	295
10.3	Schematic of LUIMA-Annotate (Grabmair et al., 2015)	301
10.4	LUIMA annotation environment	308

	<i>Illustrations</i>	xix
10.5	Annotation with GATE Teamware of factors and components in a trade secret case (Wyner and Peters, 2010)	309
10.6	WebAnno annotation of trade secret factors in the <i>Mason</i> case (see Yimam et al., 2013)	310
11.1	LUIMA pipeline architecture (Grabmair et al., 2015)	317
11.2	LUIMA-Search: sample query (top) and sentence entry in Lucene database index (bottom)	319
11.3	AP of <i>LUIMA Versions v. Baseline for Eleven Queries and MAP</i> (Grabmair et al., 2015)	325
11.4	NDCG of <i>LUIMA Versions v. Baseline for Eleven Queries and Average NDCG</i> (Grabmair et al., 2015)	326
11.5	Queries for cases with propositions playing particular argument roles (from the <i>V/IP Domain</i> ). Bold-faced terms represent existing sentence-level argument-role types. Italicized terms represent legal rule requirements	330
11.6	Queries for legal factors and argument roles (from the <i>trade secret domain</i> ). Bold-faced terms represent existing sentence-level argument-role types plus additional legal factor and value-related types (see Section 11.5.3). Italicized terms represent legal rule requirements. Underlined italicized terms represent legal policies or values	331
11.7	New argument-role sentence-level types	333
11.8	Argument mining for the <i>Mason</i> case opinion. Annotations (with WebAnno) are: trade secret misappropriation legal factors, core LUIMA sentence types, proposition/premise or proposition/conclusion, and ARGUMENT SCHEMES	337
11.9	Query input data scheme. Nodes represent successive levels of a DLF-style rule tree and reasoning chain. Here user seeks cases with evidence and an evidence-based finding that “MMR vaccine can cause intractable seizure disorder and death,” in connection with a legal ruling on the “medical theory causally connects” requirement of the <i>Althen</i> rule on causation-in-fact	341
11.10	General (left) and domain-specific (right) statutory types	344
11.11	Statutory conceptual queries from the <i>Public Health Emergency</i> domain	345
11.12	Texas LENA Statutory Network (Ashley et al., 2014)	348
12.1	Two templates for targeted legal hypotheses	357
12.2	Sample legal hypotheses CCLAs should target (from the <i>trade secret misappropriation domain</i> )	358
12.3	Sample legal hypotheses CCLAs should target (from the <i>V/IP domain</i> ). <i>Issues</i> in italics.	358
12.4	RST-Tree for <i>Mason</i> excerpt showing some attribution information	371
12.5	Annotation environment for first-year law students	376



## Tables

2.1	From regulatory texts to business rules to annotations of business process (see Figure 2.7) to predicate logic forms (Koetter et al., 2014, p. 220)	61
3.1	Trade secret Factors (Aleven, 1997)	91
5.1	Legal factors and precedents regarding <i>undue hardship</i> (see Gordon, 2008b, c; Gordon and Walton, 2009; Walton and Gordon, 2009)	137
5.2	Some proof standards in Carneades (Gordon and Walton, 2006) and legal counterparts (Weiss, 2003; Feller, 2015)	146
5.3	Can CMLAs serve as a bridge between legal texts and answers humans seek?	165
6.1	Three ontology frames for legal norm, concept, and legal act with slot fillers for library regulation (see Van Kralingen et al., 1999, pp. 1135–8, 1150–3)	177
6.2	Sample Socratic legal dialogue in a microworld with argument moves ( <i>P</i> = Plaintiff, <i>D</i> = Defendant) (Ashley, 2009a, 2011)	186
6.3	Cases in Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant). Factor abbreviations are defined below in Table 6.4 (Ashley, 2009a, 2011)	189
6.4	Factors and policies in Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant) (Ashley, 2009a, 2011)	193
6.5	Proposed tests in Property-Interests-in-Quarry Microworld ( <i>P</i> = Plaintiff, <i>D</i> = Defendant) (Ashley, 2009a, 2011)	195
6.6	Hierarchical LUIMA type system: Sentence Level, Formulation, Mention, and Term Types	205
6.7	Hierarchical LUIMA Type System: Sentence level types	206
7.1	Sample inverted index	215

8.1	Confusion matrix for three classes of sentence roles (top). Three confusion tables below show total true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) for each class	239
9.1	Problems for ML vs. KE approaches to statutory provision classification (de Maat et al., 2010)	267
10.1	Some argument schemes annotated automatically (Feng and Hirst, 2011)	293
10.2	Sentence classification performance measurements (best values printed in boldface) (Grabmair et al., 2015)	305
11.1	Eleven queries submitted to CLIR system (Grabmair et al., 2015)	318
11.2	Presuppositional information in two legal claim domains	336
11.3	Structured query translated from “Finding or conclusion that MMR vaccine causes intractable seizure disorder”	340
12.1	Examples of DLF evidentiary reasons in Special Masters’ decisions (left) and possible underlying policies or principles (right)	367

## Acknowledgments

The University of Pittsburgh School of Law provided summer Dean's Scholarships that supported writing this book. Notes for and drafts of this book evolved over the course of teaching in the University of Bologna Erasmus Mundus doctoral program in Law, Science, and Technology, an opportunity for which I thank Professor Monica Palmirani. Vern Walker, Jaromir Savelka, and Thomas Gordon read prior drafts and provided helpful suggestions for which I thank them. I am especially grateful to my former Ph.D. student and continuing research colleague, Matthias Grabmair, for his careful reading and many thoughtful suggestions. Matthias's work on legal text analytics, prediction and case-based argumentation and his and Jaromir's work applying machine learning to statutes convinced me that it was time to write this book. Advising, collaborating with, and learning from Matthias have been some of the great joys of my professional life as a teacher. I would never have had a professional life as a teacher, and I would never have completed this book, without my wife Alida's constant love and support. Our daughter Alexandra, who keeps us smiling as we toil away with our research and writing, helped me select the cover art.