Georgia State University Law Review

Volume 35
Issue 4 *Summer 2019*

Article 3

6-1-2019

Automatically Extracting Meaning From Legal Texts: Opportunities and Challenges

Kevin D. Ashley
University of Pittsburgh School of Law, ashley@pitt.edu

AUTOMATICALLY EXTRACTING MEANING FROM LEGAL TEXTS: OPPORTUNITIES AND CHALLENGES

Kevin D. Ashley*

INTRODUCTION

Legal text analytics are computational techniques that apply natural language processing (NLP), machine learning (ML), and other methods to automatically extract meanings or semantics from text archives of legal case decisions, contracts, or statutes.[1] Argument mining focuses on text-analytic discovery of argument-related information in case corpora, including premises and conclusions, argument and counter-argument relationships, and sentences that play certain roles in legal arguments and decisions.[2]

By identifying argument-related and other semantic information in legal texts, new applications can improve legal information retrieval by helping to match document structure, concepts, and argument roles with aspects of the problems users seek to solve. Eventually, the extracted information could connect artificial-intelligence (AI) models of legal reasoning and argument directly with legal texts to predict and explain case outcomes.

AI is a subarea of computer science in which researchers attempt to design computer programs to behave in a manner that we call intelligent when humans perform in the same way.[3] To put it differently, researchers build computational models of intelligent

* Professor or Law and Intelligent Systems, University of Pittsburgh School of Law

1. Michael Simon, Alving F. Lindsay, Loly Sosa & Paige Comparato, Lola v. Skadden *and the Automation of the Legal Profession*, 20 YALE J.L. & TECH. 234, 253 (2018).

2. Henning Wachsmuth, Khalid Al-Khatib & Benno Stein, *Using Argument Mining to Access the Argumentation Quality of Essays*, PROC. OF COLING 2016, THE 26TH INT'L CONF. ON COMPUTATIONAL LINGUISTICS: TECH. PAPERS 1680, 1680 (Dec. 2016).

3. J. MCCARTHY, M. L. MINSKY, N. ROCHESTER & C. E. SHANNON, A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE 2 (1955). A Dartmouth research proposal in 1955 that, notably, was co-authored by Marvin Minsky, was one of the first pieces to coin the use of the term "artificial intelligence." *Id.* at 1.

behavior. In the field of artificial intelligence and law (AI and law), they build computational models of legal-reasoning behaviors. That is why text analytics is so exciting in the AI-and-law field. For the first time, it seems plausible to connect the field's computational models of legal reasoning to the text corpora that legal professionals employ, including case decisions, statutes and regulations, and contracts.

This paper surveys three basic legal-text analytic techniques—ML, network diagrams, and question answering (QA)—and illustrates how some currently available commercial applications employ or combine them. It then examines how well the text analytic techniques can answer legal questions given some inherent limitations in the technology.

In more detail, ML refers to computer programs that use statistical means to induce or learn models from data with which they can classify a document or predict an outcome for a new case.[4] Predictive coding techniques employed in e-discovery have already introduced ML from text into law firms.[5] Network diagrams graph the relations between objects and can assist in making legal information retrieval smarter.[6] The objects may be legal cases, statutory provisions, reference concepts, or communications nodes. Finally, QA systems search large text collections to locate texts or parts of texts that directly answer a user's question.[7] IBM's Jeopardy-game-winning Watson program is, perhaps, the most famous example of a QA system.[8]

4. Bernard Marr, *How AI and Machine Learning Are Transforming Law Firms and the Legal Sector*, FORBES (May 23, 2018, 12:29 AM), https://www.forbes.com/sites/bernardmarr/2018/05/23/how-ai-and-machine-learning-are-transforming-law-firms-and-the-legal-sector/#6a397a0532c3 [https://perma.cc/5U5W-JRNP].

5. Robert Dale, *Law and Word Order: NLP in Legal Tech*, MEDIUM (Dec. 15, 2018), https://towardsdatascience.com/law-and-word-order-nlp-in-legal-tech-bd14257ebd06 [https://perma.cc/63QG-YBVL].

6. Paul Zhang et al., *Knowledge Network Based on Legal Issues*, NETWORK ANALYSIS IN LAW 21 (Radboud Winkels et al. eds., 2014).

7. DANIEL JURAFSKY & JAMES H. MARTIN, SPEECH AND LANGUAGE PROCESSING: QUESTION ANSWERING 402 (3rd ed. Draft 2018).

8. *Id.*

As noted, a variety of new legal applications employ some or all of these fundamental techniques of legal text analytics. For example, Ravn[9] and Kira[10] apply text analytics to contracts to approve routine contract language or flag unusual provisions for human review.[11] Lex Machina[12] predicts outcomes of patent and other intellectual property (IP) cases based on analysis of litigation participant-and-behavior features extracted from a corpus of IP case texts. Ravel[13] employs visual maps, citation network diagrams that graphically depict how one case cites another in connection with a legal concept.[14] CaseText's CARA[15] processes a submitted brief and identifies additional cases to cite in support of arguments in the brief based on citation networks.[16] Presumably, it uses text analytics for resolving finer grained citation links among particular paragraphs in cited cases. LENA[17] generates statutory network diagrams that provide substantive visual indices into a database of relevant statutes. Ross[18] provides a legal QA service based on IBM Watson. It accepts questions in plain English and returns answers based on texts of cases, articles, and legislation.[19]

---

9. *Artificial Intelligence*, IMANAGE, imanage.com/product/artificial-intelligence [https://perma.cc/KUJ5-9U4P] (last visited Mar. 12, 2019).

10. KIRA SYSTEMS, kirasystems.com/ [https://perma.cc/UD7D-SVKB] (last visited Mar. 12, 2019).

11. Jyoti Dabass & Bhupender Singh Dabass, *Scope of Artificial Intelligence in Law* (2018), https://www.preprints.org/manuscript/201806.0474/v1 [https://perma.cc/CAJ6-2M3X].

12. LEX MACHINA, lexmachina.com [https://perma.cc/A6A7-CBHL] (last visited Mar. 12, 2019).

13. RAVEL, ravellaw.com [https://perma.cc/ULJ3-V9WU] (last visited Mar. 12, 2019).

14. Brian Studwell, *Speeding Up Legal Research by Mapping Citation Networks*, MEDIUM (Jan. 4, 2019), https://medium.com/@bstudwell/speeding-up-legal-research-by-mapping-citation-networks-a620b128d9b1 [https://perma.cc/Z7YD-JWXW].

15. CASETEXT, casetext.com [https://perma.cc/2UR5-UZ5B] (last visited Mar. 12, 2019).

16. *Products and Features*, CASETEXTS, https://casetext.com/product [https://perma.cc/HP8D-AVUQ] (last visited Mar. 12, 2019).

17. *LENA: LEgal Network Analyzer*, PUB. HEALTH DYNAMICS LAB (Dec. 14, 2015), http://www.phdl.pitt.edu/LENA/ [https://perma.cc/684N-U3Q2].

18. Anthony Sills, *ROSS and Watson Tackle the Law*, IBM (Jan. 14, 2016), https://www.ibm.com/blogs/watson/2016/01/ross-and-watson-tackle-the-law [https://perma.cc/F43M-JNYC].

19. Cecille De Jesus, *AI Lawyer "Ross" Has Been Hired By Its First Official Law Firm*, FUTURISM (May 11, 2016), https://futurism.com/artificially-intelligent-lawyer-ross-hired-first-official-law-firm [https://perma.cc/D25X-W5EP].

This paper examines these impressive new applications of legal text analytics in automated contract review, litigation support, conceptual legal information retrieval, and legal QA against the backdrop of some pressing technological constraints. First, AI programs cannot read legal texts like lawyers can. Using statistical methods, AI can only extract some semantic information from legal texts. For example, it can use the extracted meanings to improve retrieval and ranking, but it cannot yet extract legal rules in logical form from statutory texts. Second, ML may yield answers, but it cannot explain its answers to legal questions or reason robustly about how different circumstances would affect its answers. Third, extending the capabilities of legal text analytics requires manual annotation to create more training sets of legal documents for purposes of supervised ML.

To some extent, the limitations are temporary. The questions they raise are the subjects of current research concerning the feasibility of drawing inferences from information that: (1) is implicit or distributed across documents such as contracts; (2) captures substantive strengths or weaknesses of a legal scenario; (3) requires manual annotation to teach a computer to identify; or (4) should play a role in explaining the inferences.

The paper closes with some practical strategies for dealing with these limitations. It addresses the kinds of legal-process engineering and research the legal community should undertake and underwrite to address these issues and to increase the ability of text-analytic techniques to extract semantic information, draw legal inferences, and explain them.

I. Three Basic Techniques of Legal Text Analytics

ML, network diagrams, and QA have been staple technologies in AI research for decades and have all been applied to model aspects of legal reasoning and information retrieval at various times. Only recently, however, have commercial legal applications applied them in concert to textual data yielding impressive new capabilities.[20] This section discusses, in some detail, how those techniques work.

A. Machine Learning

As noted, computer programs use statistical means to learn models from legal data. They apply the models to classify documents as instances of legal concepts or issues or to predict case outcomes. In e-discovery, for instance, the documents are classified as either relevant or not to a claim in litigation or as either subject or not to attorney–client privilege.

Two types of ML are employed in computer-assisted review of documents in e-discovery: supervised and unsupervised.[21] Supervised ML classifiers are trained to later predict class labels. In the training step, the ML algorithm takes as input chunks of text, for instance, sentences from a document or the whole document, represented as term or feature vectors and a target label.[22] A term vector represents a document in terms of its words, citations, indexing concepts, or other features.[23] The term vector is an arrow from the origin to the point representing the document in a large, dimensional space with a dimension corresponding to each term and feature in the corpus.[24] The vector's magnitude in any dimension may be a function of the frequency of a term or feature in the document and in the corpus. The target label may be a binary decision made by a human expert that a document is or is not relevant to the litigation at hand.

These labeled chunks comprise a training set; the manually classified instances are used to teach the ML classifier. With this

20. Marr, *supra* note 4.

21. Charles-Theodore Zerner & Andrew R. Lee, *Finding Needles in the Haystack*, AM. BAR ASS'N (Mar. 29, 2018), https://www.americanbar.org/groups/litigation/committees/business-torts-unfair-competition/practice/2018/tips-for-e-discovery-search-technologies-part-1 [https://perma.cc/HGH8-ADZA].

22. Raheel Shaikh, *The ABC of Machine Learning*, TOWARDS DATA SCIENCE (Nov. 14, 2018), https://towardsdatascience.com/the-abc-of-machine-learning-ea85685489ef [https://perma.cc/6LSE-JTJ7].

23. KEVIN D. ASHLEY, ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS: NEW TOOLS FOR LAW PRACTICE IN THE DIGITAL AGE 401 (2017).

24. *Id.*

training set, the program develops a statistical model that captures the correspondence between certain language features in the sentences and the target label.[25]

In the prediction step, the program applies the learned model to newly input chunks of text, from the test set, also represented as feature vectors, and predicts the label to assign to the sentence. The program can be evaluated by comparing the predicted label to a manual classification by an expert.

By contrast, unsupervised ML algorithms infer categories of similar documents without a human expert preparing a training set of manually labeled examples.[26] Instead, algorithms cluster the documents by similarity based on features of their contents or metadata.[27] Humans then determine post hoc what the members of the group share and what labels to apply, if any.[28] The aim is for the clusters to correspond to something meaningful to the task at hand, for instance, to identify all of the documents relevant to a particular issue or all of the contract provisions that are different from some norm. Often, it seems the clusters do not obviously correspond to a substantively meaningful concept; instead they reflect some inconsequential, syntactical similarity among the documents. Sometimes, however, unsupervised ML is useful for segmenting the documents into clusters as a precursor to selecting training instances for supervised learning.

In legal information retrieval, supervised ML helps to classify case decisions as instances raising a particular legal issue. LexisNexis employs ML along with rule-based and manual techniques to classify cases as sharing an issue or proposition for which the case can be cited, such as: "Thirteen-year-olds should not own a vehicle."[29] By

---

25. *Id.* at 238.

26. *What is Unsupervised Machine Learning?*, DATAROBOT https://www.datarobot.com/wiki/unsupervised-machine-learning [https://perma.cc/Y9A2-68Y9] (last visited Mar. 12, 2019).

27. *Id.*

28. ASHLEY, *supra* note 23, at 247.

29. Paul Zhang et al., *Knowledge Network Based on Legal Issues*, *in* NETWORK ANALYSIS IN LAW 21 (Radboud Winkels et al. eds., 2014).

extracting a network of similar legal issues, it assists users in retrieving other cases involving the same issues.[30]

It is interesting, and perhaps somewhat alarming, that given appropriate data, ML can predict outcomes of cases with reasonable accuracy even without accounting for a case's substantive features. The Lex Machina program, a project begun by Professor Mark Lemley and colleagues at Stanford University, initially focused on predicting patent-infringement cases.[31] It predicted outcomes of IP claims based on a corpus of all IP lawsuits in a ten-year-plus period. An early paper reported an accuracy of 64%.[32] LexisNexis subsequently acquired Lex Machina.

The program applies a statistical learning model (logistic regression) to predict outcomes of new cases based on litigation participant-and-behavior information extracted from the corpus of IP decisions.[33] It employs features of cases concerning the identity of litigation participants and their behavior, including the parties to lawsuits, attorneys and law firms, judges assigned to a case, and the districts where complaints were filed.[34] The program makes predictions based on information such as the counts of participation in past cases in any role, the past win rates of nonjudicial or district participants, and the ratio of cases assigned to a judge or district in which the plaintiffs won.[35] Analysis indicated that the identities of the judge and plaintiff's law firm contributed most to predictive accuracy, followed by the defendant's identify, the district where the case was filed, the defendant's law firm, and the defendant's attorney.

---

30. *Id.* at 21–49.
31. *Intellectual Property Thought Leader Interview With Mark Lemley*, STOUT.COM, (Sep. 1, 2016), https://www.stout.com/en/insights/article/intellectual-property-thought-leader-interview-mark-lemley [https://perma.cc/C28Q-2QPX].
32. Mihai Surdeanu, Ramesh Nallapati, George Gregory, Joshua Walker & Christopher D. Manning, *Risk Analysis for Intellectual Property Litigation*, 2011 PROC. 13TH INT'L CONF. ARTIFICIAL INTELLIGENCE & L. 116 (2011).
33. *Id.*
34. *Id.*
35. *Id.*

The authors concluded that the model appeared to be "agnostic to the merits of the case[!]"[36] The litigation participant-and-behavior features seemed to serve as a stand-in for aspects of the cases' merits. Significantly, it is technologically straightforward to extract automatically from case texts information such as names of parties, firms, and attorneys. Probably the most difficult item of information to extract concerns the outcomes of the case, sometimes hard to identify even for humans (especially first-year law students). Three IP experts coded a training set of cases for ML as to outcomes.

This begs the question of whether Lex Machina could make more accurate predictions if it took the legal merits of cases into account. Automatically extracting information about the legal or factual strengths and weaknesses of a case is a technological challenge to which we return below.

ML has been applied to predict outcomes of the Supreme Court of the United States (SCOTUS) decisions.[37] The program applies a decision-tree learning model—an extremely randomized forest of decision trees—to SCOTUS cases represented in terms of specially designed features.[38] The model correctly forecasts 70% of case outcomes and 71% of Justice-level vote outcomes over a sixty-year period.[39]

Cases are represented with features that cover information about the case, the background of the Justices and Court at that time, and historical trends. The case information includes, for example, case-origin circuit, lower-court disposition, law type, issue, issue area, petitioner, and respondent.[40] Background information includes Justice, Justice gender, Segal-Cover score, and party of appointing

36. *Id.*

37. Daniel M. Katz, Michael Bommarito & Josh Blackman, *A General Approach for Predicting the Behavior of the Supreme Court of the United States*, PLOS ONE, at 2 (April 12, 2017), https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0174698&type=printable [https://perma.cc/96UL-99WM].

38. *Id.*

39. *Id.* at 8.

40. *Id.* at 5.

president.[41] Trends include overall-historic Supreme Court, lower-court trends, current Supreme Court trends, individual Supreme Court Justice, and differences in trends.[42]

Two interesting points about these features stand out. First, like those of Lex Machina, these features do not capture the particular substantive factual features of a case. The closest they come are issue and issue area. This means that neither Lex Machina nor the SCOTUS prediction program can explain their predictions in terms of the substantive legal merits of a case. Presumably, either program could provide information on the weightiest features underlying a prediction, but those features do not correspond to substantive factual features of the case.

Second, unlike the case features in Lex Machina, the SCOTUS feature values cannot be readily extracted from the texts of the decisions. Instead, the values are prepared by political scientists or engineered by experts. The Segal-Cover score, for instance, measures a Justice's "perceived qualifications and ideology" based on expert analysis of newspaper editorials prior to confirmation. The behavioral trends and trend differences are human-engineered features. They include "tracking the ideological direction" of individual and overall Justice voting behavior. Differences in these trends "include general and issue[-]specific differences between individual [J]ustices and the balance of the Court as well as ideological differences between the Supreme Court and lower courts."[43]

B. Legal Network Diagrams

At least three types of network diagrams—that is, graphs of relations between different types of objects—apply in the legal domain depending on the type of objects linked. In a citation

41. Daniel M. Katz, Michael Bommarito II, & Josh Blackman, *Predicting the Behavior of the Supreme Court of the United States: A General Approach*, ARXIV.ORG, at 6 (2014), https://arxiv.org/pdf/1407.6333.pdf [https://perma.cc/JXX5-WQBY].

42. *Id.* at 7.

43. *Id.* at 14.

network, the connected objects may be legal cases or statutory provisions.[44] A statutory network diagram connects a set of reference concepts referred to by, and subject to, regulation across multiple statutes.[45] A social network may show communications links, such as the connections among senders and receivers of email communications.[46]

1. *Citation Networks*

Ravel makes U.S. case texts accessible in a visual map. A kind of structured citation network, it shows the intercase citation relationships of cases regarding a legal concept of interest to the user. For instance, an attorney may wish to know more about the 2010 SCOTUS decision *Citizens United v. Federal Election Commission*, which permitted corporations to make independent political expenditures.[47] When she enters the concept "campaign finance," Ravel outputs a list of cases leading to or subsequently citing *Citizens United* that are relevant to search terms, such as "campaign finance." Cases regarding campaign finance, such as *Buckley v. Valeo*,[48] are represented in the citation network as circles, whose size indicates how often the case was cited. The circles are linked by lines representing citations whose thickness represents depth of treatment, a measure of the extent to which a case is cited by or discussed in the citing opinion. The circles are distributed along an x-axis showing a chronology in years and a y-axis broken into the court-system hierarchy—that is, state courts, district courts, courts of appeals, and the Supreme Court. Alternatively, the y-axis may order the circles by relevance from the top down.

In this way, a user can trace citations from the earlier *Buckley v. Valeo* to more recent cases, including *Citizens United* and beyond, to

44. ASHLEY, *supra* note 23, at 400.
45. *Id.* at 401.
46. *Id.*
47. Citizens United v. FEC, 558 U.S. 310, 365 (2010).
48. *See, e.g.*, Buckley v. Valeo, 424 U.S. 1 (1976).

cases citing that case. When the user clicks on a circle, a textual summary of the corresponding case appears at the top of a case list.

Stanford Law graduates developed Ravel and are working with Harvard Law School Library to augment Ravel's case corpus.[49] Ravel offers fee-based analytical services that focus on judicial history. These services include pointing out cases that a particular judge found persuasive in the past; presumably, these are cases that the judge has cited and with whose results the judge has ruled consistently. They also include pointing out rules and specific language the judge has favored and commonly cited. This suggests that Ravel has extracted information both from the text near the citing case's citation to a prior case and from the cited case that indicates the reason for the citation and its connection to the concept of interest.

CaseText's CARA also provides litigation support with citation networks. When a user submits a brief, a written memorandum of law, CARA identifies and summarizes additional cases to cite in support of arguments in the brief. This also suggests that CARA uses text analytics to glean more information about the citation links between citing and cited cases, perhaps identifying topics of paragraphs and information about why a case is cited. The powerful combination of ML for analyzing texts and citation network diagrams yields more substantive information about citation links.

2. *Statutory Networks*

Statutory network diagrams show relations among entities referred to by, and subject to, particular kinds of regulation across multiple statutes and jurisdictions. For example, Figure 1 shows a graph of circular nodes, each representing a type of agent in the public health system.[50] They are connected by arrows, each representing an

49. Jeff John Roberts, *Harvard Law Just Released 6.5 Million Court Decisions Online*, FORTUNE (Oct. 29, 2018), http://fortune.com/2018/10/29/harvard-law-caselaw/ [https://perma.cc/SR2V-V2TR].

50. UNIV. OF PITTSBURGH GRADUATE SCH. OF PUB. HEALTH, THE LEGAL NETWORK ANALYZER (LENA) AND THE EMERGENCY LAW DATABASE (ELDB) USER GUIDES 5 (2015), https://www.phasys.pitt.edu/pdf/PHASYS_User_Guides_LENA-Database_v4.pdf

interaction prescribed by law; here, the arrows represent a set of California statutes dealing with disease epidemics.[51] Each arrow indicates that a statute directs one agent to perform a particular task with respect to another agent. Typically, the tasks involve some kind of communication for purposes of preparedness, response, or recovery in connection with infectious-disease surveillance. The direction of the arrow denotes which is the active agent and which the receiving agent—that is, the agent with respect to whom the action is taken. Unilateral legal directives—laws directing one agent to perform a function with a partner agent—are one color (blue).[52] Bidirectional legal directives—acting agent is directed to perform functions with a partner agent and vice versa—are another color (red).[53] The thickness of an arrow represents the strength of connection between the two agents. Thicker ties denote more legal directives requiring an interaction.

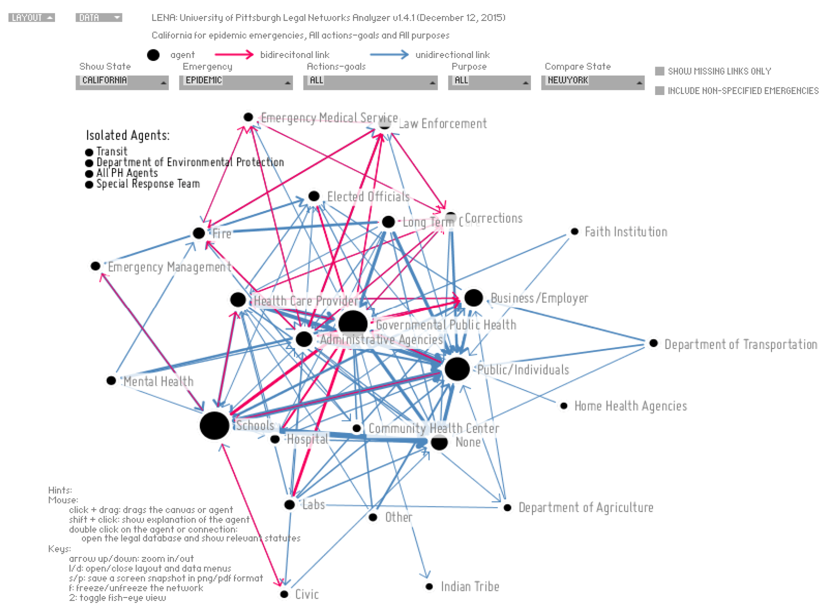
[https://perma.cc/JGRC-ZRT9] [hereinafter LENA AND ELDB USER GUIDES].

51. *Id.*

52. *Id.* at 5.

53. *Id.*

Figure 1: Statutory Network Diagram for Epidemic Surveillance in California



Each arrow is related to the set of statutes that direct a task between the two agents. The diagram's arrows thus serve as a kind of visual index into a database of statutes that direct interactions between two agents. A web-based program called LENA creates the statutory network diagrams automatically using a generic graph-layout program and a database of coded statutory data from the Emergency Law Database (ELDB).[54] The Center for Public Health Practice at the University of Pittsburgh Graduate School of Public Health developed LENA and the ELDB.[55]

More specifically, the LENA network diagram in Figure 1 shows the legally directed network of agents under California law for

54. *Id.* at 2.
55. *Id.*

epidemic emergencies involving infectious diseases.[56] The size of a node is proportional to how central an agent is in the network, a combination of its outgoing edges as an acting agent and incoming edges as a receiving agent. For instance, in the figure, Governmental Public Health and Schools are central agents.[57] An isolated agent in a network is one for which there are no legally directed functions—for example, Transit.

The LENA application can superimpose two states' statutory network diagrams, making it easy to compare those states' legally directed networks for epidemic emergencies. The nodes in these diagrams are the types of public-health-system actors and partners directed by law in both states, for example, California and New York. Different colored links signify relationships present in both states, those present in California but not in New York, and those present in New York but not in California. The visual differences in the diagrams can suggest hypotheses for public health professionals to investigate, such as that LENA's ELDB for one state may be missing some relevant statutes, or more interestingly that one state's legislature may have missed an opportunity to adopt certain provisions for dealing with epidemic emergencies that another state's legislature has found propitious.

The LENA researchers manually annotated eleven states' statutes in the ELDB to enable the network diagrams to be generated automatically.[58] In each statute of interest, certain actor agents in a state's public health system are directed with some level of prescription to perform certain actions with respect to certain receiver agents to achieve specified goals and purposes with respect to certain emergency or disaster types in a particular timeframe under certain conditions. The researchers developed a coding scheme to capture this information and used it to annotate the following coding

56. *See generally* LENA AND ELDB USER GUIDES, *supra* note 50.

57. *Id.* at 5.

58. *Prototypes*, *Research*, *Public Health Adaptive Systems Studies*, U. PITTSBURGH GRADUATE SCH. PUB. HEALTH, https://www.phasys.pitt.edu/research/prototypes.html [https://perma.cc/JM69-35DG] (last visited Jan. 31, 2019).

concepts: citation, public health agent—actor, prescription, action, goal, purpose, emergency or disaster type, public health agent—receiver, conditions, and timeframe.[59]

Because manually encoding the eleven states' statutes was time-consuming and expensive, the School of Public Health hypothesized that ML could automatically encode the statutes using manually encoded provisions as a training set.[60] A team at the University of Pittsburgh Intelligent Systems Program undertook a series of experiments to assess that hypothesis.[61] They demonstrated some success in applying ML to encode unseen statutes.[62] The results were limited by the fact that the manual encodings, recorded in Excel tables, were disconnected from the locations in the texts justifying the encodings.[63] The utility of inline annotation is discussed below. The team also demonstrated that an active-learning approach, similar to that applied in some predictive-coding approaches in e-discovery, was effective[64] and that ML models based on one state's encoded statutes could jumpstart the learning of models for other states' data.[65]

3. *Social Networks*

In legal contexts, social networks may represent communication relations among entities, such as connections among senders and receivers of corporate e-mails. As such, they may represent who communicated with whom about what and when, information that

59. LENA AND ELDB USER GUIDES, *supra* note 50, at 12.

60. Matthias Grabmair, Kevin D. Ashley, Rebecca Hwa, & Patricia M. Sweeney, *Toward Extracting Information from Public Health Statutes Using Text Classification and Machine Learning*, 235 FRONTIERS ARTIFICIAL INTELLIGENCE & APPLICATIONS 73, 73 (2011).

61. *Id.* at 74.

62. *Id.* at 79.

63. *Id.* at 80.

64. Jaromír Šavelka, Gaurav Trivedi & Kevin D. Ashley, *Applying an Interactive Machine Learning Approach to Statutory Analysis*, 279 FRONTIERS ARTIFICIAL INTELLIGENCE & APPLICATIONS 101, 107 (2015).

65. Jaromír Šavelka & Kevin D. Ashley, *Transfer of Predictive Models for Classification of Statutory Texts in Multi-Jurisdictional Settings*, 15 INT'L CONF. ON ARTIFICIAL INTELLIGENCE & L. 216, 216 (2015).

can help identify relevant communications for purposes of e-discovery. Expert e-discovery consultants use social networks of e-mails and other documents to identify senders and receivers who may harbor additional sources of data that a party should make available.[66]

The networks also indicate who has information about particular transactions that may be of interest to litigators. For instance, Figure 2 shows a network depicting all 330 e-mails in the Enron e-mail dataset responsive to the query "Blockbuster."[67] This dataset, which was produced in the giant corporate-fraud litigation involving Enron Corporation, was the subject matter of a TREC Legal Track competition.[68] In 2000, Enron and Blockbuster Corporations announced a strategic alliance, only to call it off in March 2001.[69] Hans Henseler constructed a network for this collection.[70] For each responsive e-mail, he created pairs of e-mail addresses based on the "from," "to," or "cc" slot fillers.[71] The resulting network consisted of over 1,000 directed edges with nearly 750 unique e-mail addresses as vertices.[72] He decomposed the network into islands that were disconnected from other parts of the network, filtered out any nodes that lacked outgoing links (sixty-five nodes), computed the centrality of nodes using the PageRank algorithm to measure the importance of the nodes representing e-mail senders and receivers, and weighted the edges by the number of messages.[73]

Litigators planning depositions might use the result, shown in the figure,[74] as an indication of which persons were likely to have the

---

66. Personal Communication from Eric Mandel, E-Discovery Consultant, Driven Inc. (Sept. 28, 2018) (on file with author).

67. Hans Henseler, *Network-Based Filtering for Large Email Collections in E-Discovery*, 18 ARTIFICIAL INTELLIGENCE & L. 413, 424, 428 (2010).

68. *Id.* at 414.

69. *Id.* at 423.

70. *Id.* at 424.

71. *Id.* at 419.

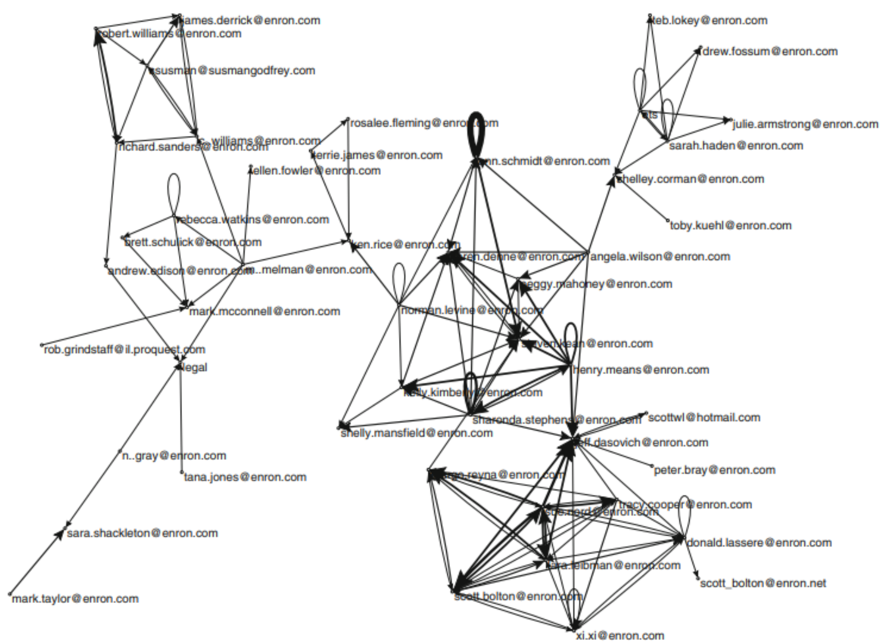
72. *Id.* at 424.

73. Henseler, *supra* note 67, at 425.

74. *Id.* at 428.

most information about the Blockbuster matter and should be sure to be deposed.

Figure 2: Reduced network for the 'Blockbuster' query with line width indicating number of e-mails between nodes in the network.



C. Legal QA

As noted, QA systems search large text collections to locate documents, short phrases, or sentences that directly answer a user's question. Ross is perhaps the best known example of a legal QA service based on IBM Watson.[75] It accepts questions in plain English such as: "If an employee has not been meeting sales targets and has not been able to complete the essentials of their employment can they be terminated without notice?"[76] It then returns an answer, citations,

75. ASHLEY, *supra* note 23, at 351.
76. *Id.*

suggested readings, and updates.[77] A team of law students at the University of Toronto developed the prototype, took second place in an IBM-hosted contest, and attracted the attention of and underwriting from IBM.[78]

For instance, Ross cites a Canadian case, *Regina v. Arthurs*, 2 O.R. 49 (1967), reports 94% confidence in the case's responsiveness, points to relevant passages in the legal text, and summarizes the decision:

> If an employee has been guilty of serious misconduct, habitual neglect of duty, incompetence, or conduct incompatible with his duties, or prejudicial to the employer's business, or if he has been guilty of willful disobedience to the employer's orders in a matter of substance, the law recognizes the employer's right summarily to dismiss the delinquent employee.[79]

Although the summary suggests that the case is not exactly on point, it seems to come close.

Ross learns from user feedback. When it returns short-text answers to a new query based on the texts of cases, articles, or legislation, its answer is followed by a request for users' feedback: "Press thumbs up if the response is accurate," or, "Press thumbs down for another response."[80] The user's feedback updates Ross's confidence in the responsiveness of its answer to a user's version of a question.

From training sets of QA pairs, Ross's ML model also learns how to assess the likelihood that it understands a user's question. Questions can be phrased in many ways, for example: "Under what circumstances can an employee be fired without warning?" The system needs to be able to recognize if the user has asked a version of a question that it knows how to answer. Experts provided legal-

77. *Id.*
78. *Id.*
79. *Id.*
80. *Id.* at 351–52.

practice questions in natural language for which a paragraph is the correct answer.

With this training set, the system learns weights associated with features of the training instances to distinguish positive or negative instances of a question. The learned weights inform the system's level of certainty that it does understand the user's question. Section II.C, below, further discusses Ross.

II. Some Limitations Affecting Legal Text Analytics

Although legal text analytics power commercial tools for automated contract review, litigation support, conceptual legal information retrieval, and legal QA, they are still subject to some major limitations concerning their inability to read or to explain their answers and their dependence on manually annotated training sets.

A. Inability to Read

Computer programs cannot yet read legal texts like lawyers can.[81] Attorneys bring a wealth of background knowledge to the task, including not only legal expertise but also common-sense knowledge about the world, human psychology, and the regulated domains at issue. AI research has not yet developed techniques for representing or applying that background knowledge.[82] Instead, using statistical methods, text-analytic programs can only extract some semantic information from legal texts.[83] These extracted meanings can be very useful. Applications can use them to improve information retrieval and ranking. There are, however, many things related to reading that these applications cannot yet do. Despite decades of attempts, AI-and-law programs cannot yet read statute texts and extract legal rules in logical form except in very limited domains, such as building

81. ASHLEY, *supra* note 23, at 13.

82. Steve Lohr, *A.I. is Doing Legal Work. But it Won't Replace Lawyers, Yet*, N.Y. TIMES (Mar. 19, 2017), https://www.nytimes.com/2017/03/19/technology/lawyers-artificial-intelligence.html [https://perma.cc/Q8Z6-SR78?type=image].

83. ASHLEY, *supra* note 23, at 11.

regulations with clearly identifiable parameters such as minimum dimensions of regulated courtyards or alleyways.[84]

In particular, computer programs cannot read contracts the way that attorneys do.[85] Automated contract review has made great strides. It applies text analytics to contract texts, semiautomates review for routine contract approval, refers apparently unusual provisions for human review, and highlights, for the reviewers, parts of texts raising apparent issues. Ravn and Kira are two of these programs.[86] They cluster contracts by topics and identify language in common, types of provisions, parameter values associated with dates, times, or dollar amounts, and recognize what is and is not boilerplate.[87] Based on this information, the applications can compare a contract's text with other contracts in a corpus to identify similarities or differences between the contract's provisions and the same type of provisions in an organization's other contracts or (dis)approved contract language. Conceivably, expert system rules could be applied given the parameter values and other extracted information to automatically approve a contract or not or earmark it for managerial review.

These tools work well on contracts with uniform language, but the tasks that lawyers face in contract analysis and due diligence often involve more complex inferences about contracts with nonuniform language. Due diligence often involves investigating a proposed transaction's assumptions and risks. A planned corporate acquisition or an investment in a legal claim for patent infringement presents risks that depend in part on the content of the target's contracts or related prior-art patents. Clearly, the task involves contract review, and automated contract review could be helpful. However, in due diligence searches, many contracts may affect the proposed

84. *Id.* at 13.

85. Lohr, *supra* note 82; *see also* Tom Simonite, *AI Beat Humans at Reading! Maybe Not*, WIRED (Jan. 18, 2018), https://www.wired.com/story/ai-beat-humans-at-reading-maybe-not/?mbid=email_onsiteshare%22 [https://perma.cc/7DV2-R8BF].

86. *Artificial Intelligence*, *supra* note 9; KIRA SYSTEMS, *supra* note 10.

87. *Artificial Intelligence*, *supra* note 9; KIRA SYSTEMS, *supra* note 10.

transaction. To what extent can text analytics enable automated contract review in a due diligence setting to extend beyond the standard structural features, provision types, and contract parameters and address information distributed across different parts of a contract or across multiple contracts? Human attorneys performing due diligence may need to make indirect inferences from information that is only implicit in the documents. Unfortunately, text analytics cannot yet extract information implicit in the texts, at least not without more knowledge and a computational model of the planned transaction, its potential risks, and how previous contractual commitments may affect them.[88]

These are not limitations for human attorneys reading the contracts in due diligence, although they may well tire of the task. Text analytics can help to flag contracts and issues that require human attention. They will not replace attorneys, however. In a nontrivial sense, contract analysis tools are illiterate, and no one would hire an illiterate attorney to perform due diligence. Instead, these tools will shift attorneys to more supervisory roles, responsible for assimilating and making sense of the information extracted automatically from multiple contracts and for drawing reasonable inferences from this information, including indirect inferences based on the human attorneys' legal expertise.

B. Inability to Explain

Although QA based on text analytics retrieves a text or text part that appears to directly answer a user's question, there may be subtle differences between the question, problem, or scenario that the author of the text was answering and the one confronting the user seeking an answer.

The jurisdiction, the applicable statute, the version of the statute applicable at a given time, or a key fact are just some of the things that may differ. Even if a QA system could be made aware of the

88. ASHLEY, *supra* note 23, at 378.

differences, it has no means for adjusting its answer except to search for another text or text part that appears to match the detailed circumstances.[89] For this, a QA system would need a computational model of the legal domain of interest with which it could reason about the appropriateness of the answer given the circumstances.

In a related way, because the QA system does not have a mechanism for reasoning about the answer, it lacks a mechanism for explaining its answer in the way that a lawyer or client might expect.[90] Of course, the retrieved text might explain the answer, in which case pointing the user to the text may also point the user to a relevant explanation. The QA system has no way to understand the explanation or even to know that it is an explanation.

To the extent that the QA system employs ML, it has selected the text based on features whose weights it has learned from training examples of QA pairs and refined based on user feedback—for instance, thumbs up or thumbs down.[91] The features may not correspond to the sort of concepts normally employed in a legal answer, and the weights may be distributed across the nodes of a neural network in a manner difficult to examine or decode. In general, ML can do an effective job of classifying texts as relevant or not but depending on the methods employed, may simply not have the information required to explain its classifications or predictions in a useful way. In any event, lacking a representation of background legal knowledge or a computational model of legal reasoning, it cannot reason robustly about how different circumstances would affect its answers. Part III, below, discusses possible approaches to deal with this inability to explain.

C. Need for Manual Annotation

As noted, supervised learning how to classify texts and parts of texts by types of semantic information requires a training set of

89. *Id.* at 352.
90. *See id.* at 352–53.
91. *Id.*

positive and negative instances of those types.[92] As a result, for text analytics and argument mining to advance, there is a growing need for legal texts that have been manually annotated with instances of the types so that an annotation pipeline can learn automatically to assign semantics to regions of text. These types include argument schemes, sentence roles in arguments, and fact patterns that strengthen or weaken particular types of claims.

As explained above, unsupervised ML from legal text collections—such as clustering—is feasible and does not require manual annotation.[93] So far, at least, it has not achieved sufficiently fine-grained clustering to be successful for automating annotation.

This raises the question of who will annotate legal texts. Crowdsourcing is a possibility; Mechanical Turk workers have annotated syntactic and certain semantic information in texts, but annotating legal semantic information requires some level of legal expertise.[94] Some interesting annotation projects have skirted the need for legal expertise by decomposing annotation tasks into well-defined subtasks simple enough for nonspecialists to perform.[95] Travis Breaux and Florian Schaub demonstrated the feasibility of crowdsourcing for annotating legal requirements in consumer-oriented privacy policies, such as: "We may *collect* or *receive* information from other sources including (i) other Zynga users who choose to *upload* their e[-]mail contacts; and (ii) third[-]party information providers."[96] Each task focused on a different target, such as identifying action verbs, types of information, sources, targets, and purposes.[97] Annotators with no legal training used an online interface to view text excerpts from privacy policies, select

92. *Id.*

93. Shaikh, *supra* note 22.

94. Robert Ambrogi, *New Legal Research Site Combines Case Law with Crowdsourcing*, LAWSITES (Jul. 26, 2013), https://www.lawsitesblog.com/2013/07/new-legal-research-site-combines-case-law-with-crowdsourcing.html [https://perma.cc/G7XR-77TQ].

95. Travis D. Breaux & Florian Schaub, *Scaling Requirements Extraction to the Crowd: Experiments with Privacy Policies*, 22 INST. ELECTRICAL & ELECTRONICS ENGINEERS INT'L REQUIREMENTS ENGINEERING CONF. 163, 163 (2014).

96. *Id.* at 166 (emphasis added).

97. *Id.* at 168–69.

and highlight phrases, encode phrases as instances of particular concepts by pressing concept keys, and highlight action verbs, such as "collect," "receive," and "upload," and relate them to the corresponding concepts of interest.[98] Online annotation environments can assist by performing some identification tasks automatically; for example, Breaux's tool used NLP to identify modal verbs.[99] Other environments can segment sentences and identify high-level parts of cases or judicial conclusions. Breaux and Schaub provided empirical evidence that crowds could successfully perform sentence- and phrase-level coding and that decomposing the workflow for coding simpler components resulted in "an acceptable aggregate response at a reduced overall cost."[100]

The users of legal applications can also annotate legal texts for ML. Having found that annotation is expensive, the ROSS Group released a free platform to which users can upload briefs for processing similar to that of Casetext Cara.[101] The platform parses the brief, analyzes it to determine whether the cited legal authorities are sound, and provides feedback on the quality of the legal analysis.[102] In exchange for this service, the platform asks users to annotate data in a kind of expert crowdsourcing activity.[103] The annotations involve highlighting the decision, key facts, and various concepts.[104] Ross provides a service that the user community needs and generates ML annotations in return, a win-win situation.[105]

Law students may annotate legal decisions as part of their studies. Annotation tasks could draw law students' attention to key aspects of the reasoning in a legal case and help students to learn to read legal

98. *Id.*
99. *Id.* at 168.
100. *Id.* at 171.
101. *Meet Eva*, ROSS INTELLIGENCE, https://eva.rossintelligence.com/#/login [https://perma.cc/X29Q-PCR7] (last visited Fed. 7, 2019).
102. *Id.*
103. *Id.*
104. *Id.*
105. *ROSS Intelligence Releases New Cutting-Edge AI, Completely for Free*, ROSS INTELLIGENCE BLOG (Jan. 29, 2018), https://blog.rossintelligence.com/post/ross-new-coworker-eva [https://perma.cc/EAG4-E9A5].

cases and statutes, one of the goals of the first year of legal instruction in the United States.[106] Cases can be annotated in terms of functional features, such as parts of decisions (introduction, factual background, and analysis) or agent identification in legal decisions. General structural features of the arguments can be marked up, such as premises and conclusions, argument relationships such as argument/sub-argument or argument/counterargument, and various argumentation schemes such as arguing by analogy. In addition, the roles that sentences play in legal argument can be annotated, including stating a legal rule, expressing a judge's holding that a rule requirement has or has not been satisfied, reporting a finding of fact, describing evidence, and reporting judges' conclusions as to issues addressed. Finally, substantive features of particular legal domains can be annotated, for example, legal factors or patterns of fact that strengthen or weaken a side's position on a claim.

For law students to learn via annotation, a convenient web-based mark-up environment is required, one that is usable on their tablet computers or laptops and that makes annotation as convenient as highlighting texts online. Law students are already inveterate "highlighters."[107] With a convenient annotation environment, they could highlight legal texts in different colors corresponding to types and produce useful data with which ML programs can annotate texts automatically. The process would sensitize them to the various functions, structures, roles, and substantive features in legal argument and give them practice in recognizing them.

Researchers at the University of Pittsburgh have developed an annotation environment called Gloss;[108] law students have begun to

---

106. *See generally* Adam Wyner, Wim Peters & Daniel Katz, *A Case Study on Legal Case Annotation*, 259 FRONTIERS ARTIFICIAL INTELLIGENCE & APPLICATIONS 165 (2013).

107. *How to Write a Case Brief for Law School: Excerpt Reproduced from* Introduction to the Study of Law: Cases and Materials, LEXISNEXIS, https://www.lexisnexis.com/en-us/lawschool/pre-law/how-to-brief-a-case.page [https://perma.cc/Z79K-XVF5] (last visited Feb. 7, 2019).

108. *Legal Glossator*, ENCYCLOPEDIA BRITANNICA, https://www.britannica.com/topic/legal-glossator (last visited Feb. 7, 2019) [https://perma.cc/8YYL-Z74A]. Gloss is named in honor of the glossatori, 11th and 12th century scholars at the University of Bologna who applied marginal or interlinear annotations (glossae) to study and teach about Justinian's 6th-century digest of Roman law. *Id.* Students of law have been annotating for a long time!

use it to annotate high-level parts of legal decisions and courts' conclusions concerning whatever issues they address.[109] Although the former is easy, finding courts' conclusions can be challenging, especially for beginning law students.

In conducting an error analysis of Gloss's automated annotation of conclusions, we found some examples that the system missed, such as: "Under these circumstances we cannot say that the trial court's finding that both Mills and Northrop understood the data to be confidential was 'clearly erroneous.'"[110] We also found some examples where Gloss predicted conclusions that the human annotators missed.[111]

Figure 3: Examples of Gloss's OVERPREDICTEDs (i.e., predicted conclusion sentences human annotator missed)

> 1. Upon the basis of this evidentiary record the Court hereby finds the following facts specially and states separately its conclusions of law thereon.
> 2. We hold that a court of equity has the power to enforce a contract against competition although the territory or period stipulated may be unreasonable, by granting an injunction restraining the respondent from competing for a reasonable time and within a reasonable area.
> 3. We conclude that this rule applies equally to both blueprints and/or drawings and customer lists because, under the facts shown, both constitute "trade secrets" within the fore mentioned definitions.
> 4. While this evidentiary record does not enable the Court to make specific findings at this juncture of the case respecting the actual monetary damage sustained by Redstone Paper Co. as a proximate result of Hughes'

109. Jaromir Savelka & Kevin Ashley, *Segmenting U.S. Court Decisions into Functional and Issue Specific Parts*, 313 FRONTIERS ARTIFICIAL INTELLIGENCE & APPLICATIONS 111, 119 (2018).
110. A. H. Emery Co. v. Marcan Prod. Corp., 389 F.2d 11, 17 (2d Cir. 1968).
111. Savelka & Ashley, *supra* note 109, at 118.

> numerous violations (as an employee of Mead) of his non-compete covenant with the plaintiff company between the dates of July 7, 1987 to the date of the conclusion of the hearing on plaintiff's application for preliminary injunction, the Court is thoroughly persuaded and finds that such monetary damage is substantial in dollar amount and will continue to grow significantly unless defendant Hughes is enjoined and restrained from continuing to commit such violations.

Although Gloss's first over-predicted example was wrong (Gloss was probably confused by the appearance of "finds" and "conclusions"), the remaining examples, and many others, were correct.[112] One could imagine a Gloss-based pedagogical environment in which students try to annotate conclusions of issues: Gloss automatically annotates and identifies some conclusions, based on prior learning, that the students missed or identifies some that it thinks are conclusions but that are not and which the students can correct.

In a process like this, students would improve at identifying conclusions and so could Gloss. In the near future, Gloss will monitor inter-annotator agreement—also known as inter-rater reliability—across multiple students by computing the level of agreement among students annotating the same documents. The human level of inter-annotator agreement sets an upper limit on ML's ability to learn; ML cannot successfully annotate concepts on which human annotators disagree. With practice, the students' skills and inter-annotator agreement would improve to the point where they generate good training data.

Much work needs to be done before text annotation becomes a regular element of legal education. The annotation activities need to be tailored into the legal curriculum, pedagogical materials must be prepared to guide students in the annotation process, and an environment needs to support students in discussing and reflecting

112. *Id.*

upon lessons from the annotation experience. Motivating student annotators to continue to perform at high levels of proficiency will depend on enriching students' social interactions about annotation, interjecting an element of competition or gamification, and enabling students to document what they have learned through a kind of portfolio development. Because objective scoring of annotation is key, and because students will be annotating documents for which no expert annotations exist, inter-rater reliability will need to serve as a standard.

Of course, annotation is subject to some general limitations. The information has to be expressed fairly directly in the texts for humans to be able to annotate it reliably or for pipeline techniques to annotate it automatically. In general, the annotation techniques will be ineffective if the information must be inferred indirectly, from multiple passages scattered across the text or from multiple documents. If the patterns in legal texts are too fine-grained, abstract, rare, or complex, analytic techniques will not be able to identify them well enough for automated annotation to work.[113] In addition, as we have seen, some useful features for prediction are engineered and not extractable from the texts, such as those based on behavioral trends in decisions of the Supreme Court, individual Justices, and lower courts.[114]

III. Research Questions Raised by Legal Text Analytics

The field of legal text analytics is still developing. The above limitations are not permanent stumbling blocks; they are challenges that inspire research questions of current interest.

First, how can legal information retrieval best employ text analytics to identify semantic information regarding the substantive merits of legal cases, for example, strengths and weaknesses or tradeoffs in effects on values? Second, how will legal applications explain, in terms that attorneys will understand, the applications'

113. *See id.*
114. ASHLEY, *supra* note 23, at 114.

answers to legal questions or predictions of outcomes? Third, to what extent can text analytics enable tasks like case analysis and automated contract review or due diligence to extend beyond standard structural features, provision types, and parameters? To what extent can it deal with indirect inferences from implicit information or address information distributed across different parts of a case, contract, or across multiple documents?

Current research addresses various aspects of all of these questions.[115] Researchers are attempting to extend text-processing pipelines' ability to learn to annotate more semantic information in the texts of cases and contracts.[116] This includes argument-related information in case texts to assist with practitioners' retrieval, explanation, and argumentation tasks.[117] The goal is to build on and surpass the previous efforts in extracting argumentative propositions, premises and conclusions, nested arguments,[118] arguments by example and other argument schemes,[119] the roles that sentences play in legal arguments,[120] and legal factors in domains like trade secret law.[121] This is where law students, legal-application users, and possibly Mechanical Turk workers can help researchers to annotate training sets of case texts with argument-related information with which ML programs can learn to identify the information in legal texts.

With this argument-related information, a program could annotate cases that a legal information-retrieval system retrieved in response to users' queries and use that semantic information, for example, to

115. *See, e.g.*, ASHLEY, *supra* note 23, at 350.
116. *Id*. at 202.
117. *Id.* at 203.
118. Raquel Mochales & Marie-Francine Moens, *Argumentation Mining*, 19 ARTIFICIAL INTELLIGENCE & L. 1, 3 (2011).
119. Vanessa Wei Feng & Graeme Hirst, *Classifying Arguments by Scheme.* PROC. OF THE 49TH ANN. MEETING OF THE ASS'N FOR COMPUTATIONAL LINGUISTICS 987, 987 (2011).
120. Bansal et al., *Document Ranking with Citation Information and Oversampling Sentence Classification in the LUIMA Framework*, 294 FRONTIERS ARTIFICIAL INTELLIGENCE & APPLICATIONS 33, 33 (2016).
121. Mohammad H. Falakmasir & Kevin D. Ashley, *Utilizing Vector Space Models for Identifying Legal Factors from Text*, 302 FRONTIERS ARTIFICIAL INTELLIGENCE & APPLICATIONS 183, 183 (2017).

re-rank retrieved cases applying an enriched model of relevance. The system could focus users on cases that better address the kind of problem that a user seeks to solve, whether it be to find a legal rule, a case illustrating how courts have applied a legal rule in specific facts, or examples of successful or unsuccessful legal or evidentiary arguments relating to these applications. This assumes, of course, that the system can discover from users' inputs and behavior the kind of problem that they seek to address with the retrieved materials; discovering more about a user's needs and constraints is itself a matter of current research.[122]

In addition to re-ranking, the program could also summarize the information in a manner tailored to the user's specific problem. Work on automated summarization has applied ML to human-prepared summaries to learn to extract sentences from cases that serve roles as introductions, context setting, reasoning, and conclusions.[123] Additional work is needed in applying more substantive annotations to construct summarizations that focus users on what a case offers that the user can really apply or say in making or responding to an argument. Thus, a summary could, for instance, briefly characterize examples of successful or unsuccessful legal or evidentiary arguments relating to the user's particular problem and to the argument that the user seeks to make. Computational models of legal argument developed in the AI-and -aw community would help to identify and characterize these arguments and, potentially, relate them to underlying legal value tradeoffs.[124]

IV. Some Practical Strategies Regarding Legal Text Analytics

Advances in legal text analytics present law firms with opportunities but also risks and questions. This section recommends some priorities for shaping a law firm's AI strategy and some

---

122. ASHLEY, *supra* note 23, at 339–42.
123. Mehdi Yousfi-Monod et al., *Supervised Machine Learning for Summarizing Legal Documents*, *in* 6085 LECTURE NOTES IN COMPUTER SCIENCE 51 (2010).
124. ASHLEY, *supra* note 23, at 141.

practical strategies for law firms to take advantage of the opportunities while reducing some of the risks.

First of all, it is important that a law firm realize that it needs an AI strategy. ML technology and data science have moved beyond e-discovery and now affect such diverse issues as predicting case outcomes and making lateral-hiring decisions.[125] Establishing a firm-wide AI committee that includes attorneys and IT staff members is a first step in assessing the potential impact. The committee should survey the firm's current uses of AI (for example, predictive coding in e-discovery), institutionalize habits of managing the firm's data sources (for example, digital files of its briefs, memoranda, employment data, and time on task information), and help to develop a culture of legal-process engineering. This means recognizing that the firm is both a consumer and a producer of law-related data and information, conceptualizing the paths, processes, and transformations of that data and information, and identifying how text-analytic techniques could add value.[126]

At virtually every step, someone with legal knowledge in the firm is conceptually linking some information about a client's facts, the provision of a contract or agreement, an applicable statute or regulation, or a precedent's facts to other information and drawing inferences and conclusions.[127] Today, some of those conceptual linkages can be preserved, for instance, through type annotation or adjustments of weights in a network, so that it can be reused in a sense, by making the firm's intelligent legal information or summarization systems more efficient or more effective.[128] Making

125. *See, e.g.*, LEX MACHINA, *supra* note 12. Lex Machina provides predictive information about firms and attorneys that could inform lateral-hiring decisions concerning IP litigators. *Id.*

126. *See* Nicholas Reed, *Legal Analytics, The Next Frontier: How Data-Driven Lawyer is Becoming Reality*, LEGALTECH NEWS (Jan. 15, 2019, 7:00 AM), https://www.law.com/legaltechnews/2019/01/15/legal-analytics-the-next-frontier-how-the-data-driven-lawyer-is-becoming-reality/?slreturn=20190107183039 [https://perma.cc/5AC9-PLPU].

127. *See The New Implications of Big Data on the Legal Industry*, BUS. COLLECTIVE, https://businesscollective.com/the-new-implications-of-big-data-on-the-legal-industry/index.html [https://perma.cc/6CEU-ZQU4] (last visited Apr. 19, 2019).

128. *See* Heidi Alexander, *Evernote as a Law Practice Tool*, L. TECH. TODAY, (Mar. 2, 2017) https://www.lawtechnologytoday.org/2017/03/evernote-law-practice-tool/ [https://perma.cc/Z3Q2-

these information processes explicit and redesigning them to generate value is the focus of legal-process engineering.[129]

Law firms can find help in performing legal-process engineering. There are, of course, commercial entities, but law firms can also connect with university researchers in computer science departments studying text analytics, ML, AI, and AI and law. Graduate students in these fields can perform useful services as paid interns in a legal firm or department. They are practitioners of a scientific empirical methodology; they understand how to evaluate text-analytic tools; they are familiar with the relevant terminology, metrics, software tools, and programming; and they understand the advantages and limitations of the technology. For instance, a graduate student in my lab has worked for two years at a major law firm ever since he invented, as a summer intern, a tool for anonymizing the firm's documents, a tool that the firm licenses to other firms. Now he helps the firm evaluate commercial technological offerings in e-discovery, automated contract analysis, ML, and NLP, all of which are related to his dissertation research. It would also spur academic research if firms financially supported it, either individually by entering into research subscriptions that some university departments support, or in collaboration with other firms. Today, academic researchers develop software innovations, some versions of which can be made widely available while restricted versions of which can be provided to subscribing firms, tailored to the particular needs and data of each firm.

The firm's AI committee should help the firm establish a sourcing strategy for deciding whether to employ an external vendor or to develop technology in-house. External vendors may have relevant expertise, but the firm may not have access to the source code or may lose control of its data. In addition, the firm is dependent on the vendor's representations and on the continuing availability of the

---

YVQK]; *see also* Darby Green, *Analytics Give Law Firms the Competitive Edge*, BIG L. BUS. (Aug. 5, 2018), https://biglawbusiness.com/analytics-give-law-firms-the-competitive-edge [https://perma.cc/JWS6-SN3L].

129. *See* ASHLEY, *supra* note 23, at 7.

vendor's expertise. Developing in-house makes it easier to arrange for the inevitable requirements of text-analytic tools for continual maintenance and adaptation to each new task.[130] In addition, the firm may license software to outsiders, which could provide an additional income source. The firm will have to acquire technical expertise in AI and legal text analytics and ensure that its employment structure can accommodate and reward a new class of expert personnel.

Finally, the AI committee should establish criteria for choosing and purchasing AI tools. This requires a commitment to understand the assumptions upon which the AI and ML analysis in any text-analytic tool is based and the data it uses. Firms should try out the programs on data with which the firm is familiar so that the results can inform intuitions about the program. Attorneys should understand how these systems are evaluated and what the evaluations signify. They should participate directly in analyzing the program's mistakes to discern possible causes of any systematic errors and correct them. They should also inspect the resulting ML models for features that principally impact predictions. This is easier to do with some learning models, such as decision trees, than with neural networks or support-vector machines, but it can be done, and the results are often illuminating.[131]

CONCLUSION

In sum, new legal applications combining fundamental text-analytic techniques of ML, network diagrams, and QA offer legal practitioners new tools to aid in legal practice. As in any profession, it is important for attorneys to understand the tools they use in practice, including where, how, and how well they work and what their limitations are.[132] Today's tools are subject to some limitations

130. Lohr, *supra* note 82.

131. Louis Dorard, *Machine Learning and the Law*, MEDIUM (Dec. 21, 2016), https://medium.com/louis-dorard/machine-learning-and-the-law-adf051d33253 [https://perma.cc/J7TR-6HCX].

132. Natalie Kelly, *2017 Solo and Small Firm Technology*, AM. B. ASS'N (Dec. 1, 2017),

in terms of their inability to read legal texts as lawyers do, to explain their answers as well as attorneys would expect, and to extract implicit information from texts.[133]

Given the opportunities for law firms and the challenges for researchers of automatically extracting meaning from legal texts, a lingering question for law schools is how best to prepare law students for changes in legal practice that will result. In spring 2019, I will pursue one possible answer. My co-instructor, a systems scientist at the Carnegie Mellon University Language and Technologies Institute, and I will offer a course entitled "Legal Text Analytics and AI" to a combined group of law students and computer science undergraduate and graduate students from the University of Pittsburgh and Carnegie Mellon University. The course will present a Python programming tutorial, introduce the field of AI and law, focus on formal rule- and case-based reasoning and computational models of argumentation, and cover some basics of ML and NLP. The course will focus in depth on analysis and prediction using the Supreme Court Database, fairness in ML given policies of nondiscrimination, and legal text analytics including annotation, rule- and ML-based text processing, and information retrieval. Along the way, the course will address various legal topics related to AI such as statistical argumentation in courts, legal liability of autonomous vehicles, and personal-care robotics.

In the latter third of the course, students will form mixed teams of lawyers and engineers and propose a final project on legal data analysis on which they will work collaboratively. The goal is to provide law and graduate students practice with applying basic tools and techniques of ML, practical experience in formulating and assessing research hypotheses in legal data analytics and in designing, and planning and critically evaluating legal data-analytics project work. Law students will gain experience communicating with

---

https://www.americanbar.org/groups/law_practice/publications/techreport/2017/solo_small_firm/ [https://perma.cc/S8N6-YV7V].

133. *See supra* Part II.

technical personnel and vice versa, learn the relevant metrics, perform error analysis, and learn a scientific method, which they can subsequently apply in legal-process engineering in their future practice.