ARTICLES

Law, Artificial Intelligence, and Natural Language Processing: A Funny Thing Happened on the Way to My Search Results [2020-6]

Paul D. Callister 161


Shawn G. Nevers and Julie Graves Krishnaswami 213

A Different Democratic Divide: How the Current U.S. Online Court Record System Exacerbates Inequality [2020-8]

Lisa von Wiegen and Shannon M. Oltmann 257
Statement of Ownership and Management

*Law Library Journal*, Publication No. 23-9283, is published quarterly (spring, summer, fall, winter) by the American Association of Law Libraries, 105 W. Adams St., Ste. 3300, Chicago, IL 60603-6225. Annual subscription price, $125. American Association of Law Libraries, publisher; Tom Gaylord, editor; Heather Haemker, managing editor. Periodical-class postage paid at Chicago, Illinois. Printed in the U.S.A. As a nonprofit organization authorized to mail at special rates (Section 423-12, *Domestic Mail Manual*), the purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding 12 months.

**Extent and Nature of Circulation**

(“Average” figures denote the average number of copies printed each issue during the preceding 12 months; “Actual” figures denote actual number of copies of single issue published nearest to filing date: Spring 2020, vol. 112, no. 2 issue [published October 14, 2020].) Total number of copies printed: Average, 225; Actual, 132. Paid and/or requested circulation: not applicable (i.e., no sales through dealers and carriers, street vendors, and counter sales). Mail subscription: Average, 73; Actual, 58. Free distribution by mail, carrier or other means, samples, complimentary and other free copies: Average, 40; Actual, 29. Total distribution: Average, 92; Actual, 64. Copies not distributed: office use, leftover, unaccounted, spoiled after printing: Average, 150; Actual, 150. Returns from news agents: not applicable. Total (sum previous two entries): Average, 282; Actual, 243. Percentage paid: Average, 73%; Actual, 68.82%.
Table of Contents

General Articles

Law, Artificial Intelligence, and Natural Language Processing: A Funny Thing Happened on the Way to My Search Results [2020-6]  
Paul D. Callister 161

Shawn G. Nevers 213  
Julie Graves Krishnaswami

A Different Democratic Divide: How the Current U.S. Online Court Record System Exacerbates Inequality [2020-8]  
Lisa von Wiegen 257  
Shannon M. Oltmann

Review Article

Keeping Up with New Legal Titles [2020-9]  
Susan Azyndar 275  
Susan David deMaine
LEGAL CAPITAL, 4th Edition

Bayless Manning, Late Former Dean, Stanford Law School
James J. Hanks, Jr., Partner, Venable LLP

Legal Capital is widely credited with pioneering the introduction of the balance sheet and equity solvency tests, as well as other reforms in the Model Business Corporation Act and corporation statutes in over 30 states. This edition adds new historical material, updates the statutes and caselaw on dividends and other distributions in the U.S., and compares the evolution of legal capital in countries around the world.

Visit | store.westacademic.com
Call | 800-782-1272
email | inquiries@westacademic.com

© 2009 L.L.C. Inc. www.West Academic; West; West Academic Publishing; and West Academic are trademarks of West Publishing Corporation, used under license. 101735-01
Artificial intelligence (AI), including natural language processing, may challenge the legal profession as much, if not more, than the shift from print to digital resources. We may be inevitably moving toward letting AI become our touchstone for authority or, as Robert Berring has articulated, our “cognitive authority.”

Introduction ......................................................... 162
Issues and Themes .................................................... 164
Introductory Quotations: Elaborations ............................. 164
More Immediate and Practical Concerns ............................ 168
Building a Model of an Information Retrieval System with Natural Language Processing ........................................ 173
Trigonometric Word Vectors and Cosine Similarity ..................... 173
TF-IDF or Term Frequency and Inverse Document Frequency ................ 175
Centroids and Document Similarity ..................................... 178
Queries as Vectors—An Example ..................................... 179
word2vec and Skip-gram with Negative Sampling .......................... 181
Potential Use of word2vec in Information Retrieval ...................... 182
Regression and Neural Networks ..................................... 183
Classification ...................................................... 183
Grammatical Techniques ............................................ 186
Bag of Words, word2vec, and the Performance of Current Legal Search

Engines .......................................................... 187
Westlaw Edge ...................................................... 187
Lexis Advance ..................................................... 189
Bloomberg Law .................................................... 192
Fastcase .......................................................... 193
Casetext .......................................................... 195
Ravel ............................................................. 196
Google Scholar ..................................................... 197
Caution and Summary .............................................. 198

* © Paul D. Callister, 2020. Licensed for reuse under Creative Commons Attribution 4.0 International license, https://creativecommons.org/licenses/by/4.0/ (last visited Aug. 4, 2020).
** Library Director, Leon E. Bloch Law Library, and Professor of Law, University of Missouri-Kansas City School of Law, Kansas City, Missouri. I wish to give special thanks to Pablo Arrendondo of Casetext for encouraging me to look into words as vectors. I wish to thank Susan Nevelow Mart, Dr. Peter Hook, and Kim Kirschenbaum for the assistance in editing and advice.
Word Prediction Techniques and Secondary Sources ...................... 199
Teaching Search Inquiry in the Age of Natural Language Processing ...... 203
Differences in Results ................................................. 203
Lengthening Search Inquiries ......................................... 204
Recognizing Biases of Search Algorithms ................................ 204
Recognizing Out-of-Context Matches ................................... 205
Secondary Sources and Codes ........................................... 205
Reiterative Searching ................................................... 205
Search Result Visualizations ............................................. 207
Conclusion .............................................................. 207
Appendix A .............................................................. 211
Appendix B .............................................................. 212

“Law must be stable and yet it cannot stand still.”—Roscoe Pound

“Like gods, these mathematical models were opaque, their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists.”—Cathy O’Neil

“For most of the twentieth century, the legal world had agreed to confer cognitive authority on a small set of resources. By ‘cognitive authority’ I mean the act by which one confers trust upon a source.”—Robert C. Berring

Introduction

Artificial intelligence (AI), including natural language processing, may challenge this profession, including the larger profession of law practice, as much as, if not more than, the shift from ownership of print resources to licensed digital resources. We may be inevitably moving toward AI becoming our touchstone for

---

5. Natural language processing may be defined as a computer’s “way of processing language as actually used rather than set commands.” Jones, Kalandery & Glover, supra note 4, at 5. “Natural Language Processing (NLP) is the processing and analysis of unstructured language data, essentially enabling computers to understand human language.” Id. at 8. For example, a lot of scientific information is structured and can be processed easily, but humanities and social science information (including law) is unstructured. Id.
authority or, as Robert Berring has articulated, our “cognitive authority,” but we are not there yet. Outside of the field of law, it is easy to see how this shift has already occurred. We can ask Google and Amazon Alexa about all sorts of things and get cogent answers. For example, “Who won the last Chiefs game?” More remarkable than the answers is that we trust their accuracy. These devices and the software that supports them have become part of our cognitive authority. Within law, the relationship between search engines and cognitive authority is more complex.

¶2 To serve the topic, this article will proceed as follows: first, beginning with the introductory quotations, the issues and themes will be introduced along with practical issues surrounding student research habits and problems with inconsistent research results. Next, to understand how effective natural language processing (a subset of AI) is in current legal research, I will go about building a model of a legal information retrieval system that incorporates natural language processing. I have to build my own because we do not know very much about how the proprietary systems of Westlaw, Lexis, Bloomberg, Fastcase, Ravel, and Casetext work. However, there are descriptions in information science literature and on the Internet of how systems with natural language processing actually work or could work. Then, I will compare such systems with the features and search results produced by the major vendors to illustrate the probable use of natural language processing, similar to the models. Next, the use of word prediction or type-ahead techniques in the major research services is also worth studying while considering natural language processing—particularly, how such techniques can be used to bring secondary resources to the forefront of a search. Finally, I will explore how the knowledge gained may help us to better instruct law students and attorneys in the use of the major legal information retrieval systems.

¶3 My conclusion is that the adeptness of natural language processing is uneven among the various vendors and that what we receive in search results from such systems varies widely depending on a host of unknown variables. Natural language processing has introduced new uncertainty to the law. We are a long way from idyllic AI systems that understand, let alone search, legal texts in a stable and consistent

6. Robert Berring laid out his idea of cognitive authority in Berring, supra note 3, at 1676 (“The cornerstone tools of legal information have been established as unquestioned oracles. They appeared too obvious to examine.”).

7. For this article, I am using natural language processing as a single example of AI, and will generally refer to it instead of AI. Natural language processing is not the same thing as natural language searching, although the latter tends to incorporate the former. Natural language processing has many other uses than information retrieval incorporating relevancy-based feedback. Its ultimate aim is to process language as humans do and respond in kind. See Jones, Kalantry & Glover, supra note 4.

8. While I was writing this article, an important article appeared in AALL Spectrum, coauthored by several important data scientists from various vendors and librarian Susan Nevelow Mart, that discusses how the major vendors employ search algorithms. Susan Nevelow Mart et al., Inside the Black Box of Search Algorithms, AALL Spectrum, Nov./Dec. 2019, at 11. However, the discussion is in general terms about search features, and does not include the level of detail, often mathematical, in this article—incomplete though it may be. The Spectrum article mentions TF-IDF (also discussed infra, ¶¶ 27–34), automatic query expansion, term proximity, result ranking, term proximity to key words, the importance of context to relevancy, source Authoritiveness, the aggregation of user search history, support for search commands, differentiation of search engine by type of document, machine learning algorithms, probabilistic approaches, etc. See id. at 11–15. Even with this laundry list of tools, the details of just how current information retrieval features work and are applied are left vague, which is in part due to the short nature of Spectrum articles, but may also be a result of the proprietary nature of search algorithms.
way. The use of secondary authority, such as the great legal treatises, is at risk. This may make establishment of cognitive authority more difficult. On the other hand, bestowal of cognitive authority on our new tools may be an act of faith and may even be necessary for us to function as a profession, driving us toward AI systems in the future. It may also be inevitable that only a small cadre of the profession may take advantage of what will increasingly become extremely powerful, but costly, systems for providing legal research assistance.

Issues and Themes

¶4 The objective of this section is to introduce both the theoretical issues surrounding AI and cognitive authority and the practical concerns arising from our current (and future) uses of natural language processing.

Introductory Quotations: Elaborations

¶5 The three quotations that began this article suggest three issues we face as law librarians (and which the larger law profession faces too). The first quotation from legal scholar and educator Roscoe Pound reminds us of the need for tension between stability and adaptability in the law. Yet, how will the law remain stable when, as demonstrated in this article, our tools for accessing the law provide such disparate results? The second quotation goes to the issue of opaqueness in natural language processing and the algorithms that are producing our search results. It was made by a prominent data scientist who wrote a book, *Weapons of Math Destruction*, warning of the devastating effects of unintended consequences of algorithms in everything from lending markets to education. We are unlikely to gain access to how prominent vendors tune their algorithms. They remain opaque, yet authoritative. “Authority is increasingly expressed algorithmically.” As in other domains, are there harmful and unforeseeable unintended consequences from natural language processing of legal authority and commentary? There are, primarily because of inconsistency among services and lack of access of most attorneys to all of the search tools providing such differing results.

¶6 The third quote deals with “cognitive authority,” which is a conferral of trust on a historically small set of particular legal research resources. This bestowal of authority could be on a primary source, like the *United States Code Annotated*, even though it is not the official version of the *United States Code*. The legal community

10. See infra ¶¶ 53–74.
12. Although I was able to find surveys on what attorneys use, I was unable to find information about services to which attorneys typically do not have access. See Robert Ambrogi, Survey Finds Virtual Dead Heat in Lawyers’ Use of Westlaw, LexisNexis and Fastcase, LS LawSites, Mar. 13, 2017, https://www.lawsitesblog.com/2017/03/survey-finds-virtual-dead-heat-lawyers-use-westlaw-lexisnexis-fastcase.html (last visited Aug. 4, 2020). Admittedly, my statement about lack of access to the full range of legal research services is based on anecdotal evidence. This would be worth studying in a future article.
13. See Berring, supra notes 3 and 6 and accompanying text.
can entrust cognitive authority to commentary like Nimmer on Copyright, Moore’s Federal Practice: Civil, or BNA Tax Portfolios.

¶7 The concept of cognitive authority shares affinity with other core concepts, which in past publications I have referred to as legal epistemology or the shared web of beliefs common to the legal profession.14 Ronald Deibert, from whom I adapted the concept of legal epistemology for the profession, uses the concept of social epistemology for the larger society, which is the “web-of-beliefs into which people are acculturated and through which they perceive the world around them.”15 That web-of-beliefs includes what constitutes cognitive authority.

¶8 There is a relationship between cognitive authority and legal institutions, technologies, language, and even the geopolitical environment. They affect one another as part of a holistic model of the legal information ecosphere, as illustrated by figure 1, adapted from Deibert’s ecological holistic model of media theory.16


15. See Book as Authoritative Sign, supra note 14, at 51 (citing RONALD J. DEIBERT, PARCHMENT, PRINTING, AND HYPERMEDIA: COMMUNICATION IN WORLD ORDER TRANSFORMATION 94 (1997)).

16. See Deibert, supra note 15 at 38. For a detailed interpretation of each ring and the arrow connecting past and future based on historical criteria, see Law’s Box, supra note 14, at 267–72. The model is referred to as holistic and ecological because Deibert wanted to contrast his formulation with prior versions of media theory that were determined by changes in technology—most notably from Harold Adam Innis and Marshall McLuhan. See id. at 265.
¶9 Note in figure 1 that technology and language are tied together. We can see that quite clearly with natural language processing, which is about the attempts to deal computationally with unstructured language, as is found in legal discourse.17 Language is a quintessential technology. Other examples of the relationship of language, technology, and institutions include the ancient Egyptians, who because of silent determinants of meaning within hieroglyphic and hieratic writing required a scribe class to interpret rather than just phonetically vocalize law;18 the Celtic and Icelandic bards who used meter and devices like stating the law in triads, to preserve and communicate the law;19 the gloss of medieval codices and manuscripts, which sometimes made its way into the law itself;20 the use of clay “wrappers” or seals to authenticate legal documents in Mesopotamia;21 the demotic (meaning common or democratic) alphabet of the classical Greeks used to communicate law openly on stone stele to whomever could read;22 and the use of pinpoint citation in the era of the printing press to stabilize and create a web of legal authority.23 Now the issue is potentially the natural language processing of legal texts to enable machine understanding and participation in legal dialogue. It’s not just about searching the law; it’s about understanding it.24 Natural language processing may become society’s new mediating scribe of the law—able to understand relationships of legal texts invisible to human intelligence without its aid. Such a technology would quite naturally be endowed with cognitive authority by the legal community.

¶10 Returning to figure 1, the arrow on the chart is significant. Its tail suggests that the past bestowals of cognitive authority cannot be ignored. There is not a

New technologies of communication do not generate specific social forces and/or ideas, as technological determinists would have it. Rather, they facilitate and constrain the extant social forces and ideas of a society. The hypothesized process can be likened to the interaction between species and a changing natural environment. New media environments favor certain social forces and ideas by means of a functional bias toward some and not others, much the same as natural environments determine which species prosper by “selecting” for certain physical characteristics. In other words, social forces and ideas survive differentially according to their “fitness” or match with the new media environment—a process that is both open-ended and contingent.

Deibert, supra note 15, at 36.

17. See Jones, Kalantery & Glover, supra note 4. In contrast to the social science (and law), the language of the hard sciences is considered to be structured. Id.


19. See id. at 311–19. The medieval Welsh used triads. See generally The Legal Triads of Medieval Wales (Sara Elin Roberts ed., 2007).

20. See Law’s Box, supra note 14, at 308.

21. See id. at 285–86.

22. See id. at 278. The Greek alphabet, unlike Egyptian hieroglyphics and hieratic, was purely phonetic, without silent determinatives that influenced meaning, requiring a scribe (in the case of Egyptian, but not Greek) to interpret the meaning. Compare id. at 278 with id. at 296–300. The role of language in liberating the common people from scribal and bureaucratic classes has been noted: “[T]he eminent British legal historian and diplomatist M.T. Clanchy entertained . . . criticisms in his monumental work on the evolution of English legal documents: ‘[I]t is language itself which forms mentalities, not literacy . . . . Morally and psychologically, depending on the circumstances, literacy may liberate or it may confine.’” Id. at 300 (citing M.T. Clanchy, From Memory to Written Record: England 1066–1307, at 9 (2d ed. 1993)).

23. See Book as Authoritative Sign, supra note 14, at 66–69.

24. “‘Legal research’ is not merely a search for information; it is primarily a struggle for understanding.” Michael J. Lynch, An Impossible Task but Everybody Has to Do It—Teaching Legal Research in Law Schools, 89 Law Libr. J. 415, 415 (1997). See also Nevelow Mart et al., supra note 8, at 15 (describing current application of Westlaw Edge’s natural language processing to “understand the meaning of a query”).
complete break between successive periods of cognitive authority. Each period from the past influences what is accepted in the future. For instance, Lord Coke, in writing his famous treatise, or *Institute*, or *Commentary on Littleton*, used a format suggestive of the Justinian gloss that came from the manuscript era prior to the printing press (even though Lord Coke’s treatise was published on such presses).* In more recent times, West and Lexis, including their migrations to electronic versions, have earned what I would argue is cognitive authority for access to primary authority of the law—cases, codes, and regulations—and even secondary authority for the great treatises cited in court. They are accepted almost without reservation despite often not being the official versions, in the case of primary law. However, even online services have to acknowledge the past by preserving page numbers from print sources, and in some cases making available PDF images of the print versions. The point of all these examples and the chart in figure 1 is that technologies such as language affect institutions—including the bar, courts, and legal publishers—which in turn affect (and are affected by) cognitive authority. Natural language processing will have to ground itself in the forms and functions of cognitive authority of the past—perhaps such as giving cognizance to most-cited cases, adhering to jurisdictions, performing citation analysis, building on West’s Topic and Key Number System, emphasizing cases annotated in *American Law Reports*, or any number of a hundred factors that make up the current terrain of the legal information environment.

Looking at the point of the arrow in figure 1, which represents the potential of the future, we live in a remarkable age. The whole purpose of natural language processing is to give machines more “understanding” of (or at least effectiveness with) human speech as it is actually delivered. Efforts to employ natural language processing include application to legal texts. Predictably, there will be an effect upon the legal profession’s cognitive authority, while at the same time the legal profession’s traditions related to cognitive authority will affect how natural language processing is employed and even accepted. My prediction is that if natural language processing becomes increasingly adept at answering legal questions with precision, reliance upon the great treatises and secondary sources may falter—the relationship between primary law and secondary authorities may be usurped by a new relationship based on natural language processing.* However, as this article will hopefully

---


26. *See Book as Authoritative Sign*, supra note 14, at 66–67 (in particular, note replication of Lord Coke’s first *Institutes*, based on Littleton, in the manner of glossed texts, in fig. 3.7).

27. My predictions about the replacement of secondary authority (authored by human experts) by natural language processing are not out of line with forecasts about the relationship of work done by artificial intelligence instead of humans. Just to illustrate, the World Economic Forum estimates that “the average percentage of tasks carried out by machines vs. humans will change from 29% vs. 71% in 2018 to 42% vs. 58% [by 2022].” Mary Lee Kennedy, *What Do Artificial Intelligence (AI) and Ethics of AI Mean in the Context of Research Libraries?*, in *Association of Research Libraries, Ethics of Artificial Intelligence, Research Libr. Issues* no. 299, 5 (2019), https://publications.arl.org/rli299/1 (last visited Aug. 4, 2020). Supporting this conclusion: “In a 2018 survey of AI researchers, 50% forecasted that high-level machine intelligence (HLMI) would be achieved within 45 years [and a 10% chance within 9 years]. HLMI is achieved when machines can accomplish every task better and more cheaply than human workers.” Jones, Kalantery & Glover, *supra* note 4, at 17 (citing Katja Grace et al., *When Will AI Exceed Human Performance? Evidence from AI Experts* (2017, revised May 2018), https://arxiv.org/pdf/1705.08807 (last visited Aug. 4, 2020)). In the field of law, Richard
reveal, the current state of technology, while trending away from secondary sources, is still a ways from replacing them. This seems particularly true in that current legal natural language processing has little facility with legal syntax, but depends instead upon general word proximity and co-occurrence. Of course, as shall be shown, the current state of things is more nuanced and complex than simply determining word proximities and co-occurrences.

More Immediate and Practical Concerns

¶12 Turning to practical concerns, I have long maintained that legal researchers should vary search techniques according to the types of problems they face.28 But I must confront the reality made evident by a recent survey of law students at my school. In discussing how they start their research, almost half the students agreed with this statement: “I just start typing in the search bar on Lexis, Westlaw, or Bloomberg to see what comes up.”29 This held true even though “starting with Google” was also an option on the survey (12.87% of students selected). Starting with legal commentary (something with an index) to get background on the law is far less likely (11.88%) than using a single-search box. Students can filter their results on Lexis, Westlaw, and Bloomberg “post-search” to find secondary sources, but do they bother when primary materials are displayed first and may seem relevant?

¶13 We have entered the age of the one-size (or one search box) fits-all legal inquiry, and dare I suggest that what will follow (especially, in light of vendor advertising) is that one day we will start to behave as if legal search algorithms are intelligent and capable of understanding our queries.30 This will only get truer with

Susskind revised his best-selling book, Tomorrow’s Lawyers: An Introduction to Your Future, to assert, “as our machines become increasingly capable, they will steadily eat into lawyers’ jobs. The best and the brightest human professionals will last the longest . . . .” Richard Susskind, TOMORROW’S LAWYERS: AN INTRODUCTION TO YOUR FUTURE 188 (2d ed. 2017). Those jobs may ultimately include the recognized attorney/editors and authors of the major commentaries that are part of the cognitive authority of the law.


29. The question in the 2019 survey of 101 UMKC law students was:

Which of the following statements are true about starting your legal research online? Select only one.

☐ I just start typing in the search bar on Lexis, Westlaw, or Bloomberg to see what comes up. 48.51%

☐ I tend to dive into a search of primary law (cases, statutes, etc.) as my first effort. 20.9%

☐ I like to start with legal commentary like Missouri Practice, American Law Reports, or treatises to get background on a topic before beginning my research. 11.88%

☐ I have looked up law review and journals online for my topic at the start of my research. 5.94%

☐ I start with Google to find information on point. 12.87%


30. Westlaw Edge has released WestSearch Plus, which will answer a variety of legal questions with “type-ahead functionality” and search suggestions. Patrick Yatchak, Thomson Reuters, Answer Legal Questions Faster Than Ever with West Search Plus, https://legal.thomsonreuters.com/en
each new search engine released by the vendors. *AALL Spectrum* reports, “Recently, Westlaw Edge extended those capabilities through a set of proprietary natural language algorithms that aim to ‘understand the meaning of the query’ and, when appropriate, provide answer-like results.” It is entirely possible that legal information systems will progress from search boxes to research assistants to perhaps (as others have predicted) even providing legal services. The paramount question moves from “will these systems provide stable access to the law?” to “will there even be a stable system of American law in such a world?” Relatedly, will the algorithms or AI driving such systems themselves become the profession’s basis for cognitive authority—a concept that asks whether they will be the trusted sources of American law, perhaps in part by subjugating the treatise?

Historically, in a book environment, legal research (and law) was stable by design. We all had the same digests and indexes from West and citators from Shepard’s. But for some time the very fabric of legal research inquiry (at least its initial steps) has been torn by the unpredictability of diverse search algorithms and methods for natural language processing. This concern occurs at a time when...
according to a recent study of 325 decisions in the federal courts of appeals (citing 7552 cases), only 16% of the cases cited in appellate briefs make it into the courts’ opinions.35 The single-search box raises several concerns in addition to the concurrent weakness of case citation in appellate briefs.

¶15 First, Susan Nevelow Mart has demonstrated in a seminal article, The Algorithm as Human Artifact: Implications for Legal [Re]Search, that the different online research services (Westlaw, Lexis Advance, Fastcase, Google Scholar, Ravel, and Casetext) produce significantly different results when researching case law with natural language techniques.36 Roughly 40% of the search results in each studied database were unique.37 For the top 10 results, when combining unique and relevant results, statistics varied greatly, with new services clustering at lower numbers, and Lexis and Westlaw having higher rates (Casetext 8.2%, Fastcase 13.1%, Google Scholar 14.6%, Lexis Advance 19.7%, Ravel 11.3%, and Westlaw 33.1%).38 “These algorithmic variations in worldview lead to substantial variations in the unique and relevant results each database provides. The knowledge of this variability expands the opportunities for researchers to find relevant cases that can play ‘some cognitive role in the structuring of a legal argument,’ ”39 The point is how can the law be stable, a fundamental axiom,40 when our research tools are providing “substantial variations in the unique and relevant results?”41 This is a theme that I shall demonstrate also proves true in some of the sample natural language searches done on case law for this article.

¶16 While Nevelow Mart sees the diversity in unique results as an opportunity, few attorneys have access or time to use all of the tools employed in her study (nor can clients afford it). When I asked Nevelow Mart about this issue, her answer is that the solution to having limited access to the different services is to use reiterative searching to discover all of the relevant material.42 Whether such methods are

35. Ken Bennardo & Alexa Z. Chew, Citation Stickiness, 20 J. APP. PRAC. & PROCESS 61, 82, 74–75 (2019). “In our 325-case data set, the parties cited 23,479 cases. Of those, only 16% were later cited by the courts in their opinions—or to use our nomenclature, only 16% of the cases cited in the briefs were sticky.” Id. at 84. If both parties cited the case, 38% of such opinions were cited by the appellate court. Id. The same study found that 49% of cases cited by courts had been cited by at least one party in their brief (only 21% were cited by both parties). Id. Samples were taken from each of the Circuit Courts of Appeal. Id. at 78. It would be interesting to study whether these statistics hold true in earlier periods, prior to digital search engines.

36. See Nevelow Mart, supra note 34, at 397, ¶ 16; 409, ¶ 36. Technically, Nevelow Mart conducted “key word” searches, which she distinguishes from natural language searches, the latter of which employ “grammatical structures models.” Id. at 397, ¶ 16. See infra ¶¶ 51–52 for discussion of grammatical techniques. I do not draw such a distinction between key word and natural language searching because I am writing about natural language processing, which includes both key word and “grammatical structures” searches.

37. See id. at 413, chart 1.

38. See id. at 415, chart 3. The relevancy of top 10 cases also varied with Westlaw at 67%, Lexis Advance at 57%, and the rest clustered around 40%. See id. at 414, chart 2.

39. Id. at 420, ¶ 57 (citing Stuart A. Sutton, The Role of Attorney Mental Models of Law in Case Relevance Determinations: An Exploratory Analysis, 45 J. AM. SOC’Y INFO. SCI. 186, 187 (1994)). For what Nevelow Mart means by “worldview,” see supra note 34.

40. See, e.g., POUND, supra note 1.

41. See Nevelow Mart, supra note 34, at 420, ¶ 57. Roughly 40% of search results across the different services are “unique to one database.” Id. at 390, ¶ 5.

a complete solution to the problem of diverse results among the major and minor search engines is worthy of further study. In particular, we need to test whether reiterative searching on the staples of Lexis and Westlaw and newcomers Fastcase (supplied with some state bar memberships), Casetext, Ravel, Google Scholar, and even Bloomberg will find the same seminal cases that give stability to the law. If not, the profession may be doomed to a lack of confidence in its search results. Although there may be manifold explanations for the phenomenon, given the diversity of search services, there is little wonder courts and attorneys are at odds with respect to case citations in decisions and briefs. Cognitive authority suffers in such an environment.

¶17 The second concern I have is that legal research problems that are primarily subject in nature are more readily solved, at least historically, by using different techniques (such as indexes and tables of contents with legal commentary) as opposed to known-item problems that best use search algorithms and are fact specific (“I need the case where [insert determinative facts and narrow issues]”). I have held to this fundamental belief and, in a general sense, have felt that more often than not, research involves subject problems (“I need to understand Wisconsin water law”), and thus requires techniques other than search algorithms to resolve. But do the assumptions inherent in all of this continue to hold true, if they were ever true in the first place? More concretely, is natural language processing, a form of AI, and as used by our major vendors, so good that it no longer matters what problem we face because something authoritative and relevant will always appear in the search results for any given inquiry (even prior to filtering sources)? Legal commentary or secondary authority (the classic texts the profession has relied on, say, Williston on Contracts, Nimmer on Copyright, or Moore’s Federal Practice) might become invisible with relevant primary authority appearing as defaults in search results. Furthermore, new legal search engines lack connection with much of the commentary and treatises, which are a part of the profession’s cognitive authority. Finally, natural language processing may ultimately produce outputs to search queries in forms of abstracts or summaries that replace those commentaries and treatises.

¶18 I also wonder whether because of AI and natural language processing, we are observing the death throes of the index and the end of human-intermediated access tools to legal information, such as digests and abstract services. My survey of students was not a fine enough instrument to definitively determine this phenomenon, but it does raise important questions. Furthermore, can natural language processing get us into secondary sources as well as indexes have? And where and how will statutory codes be effectively accessed? They are arranged, at least when codified, with a topical structure and have indexes. These are likely to be ignored in

43. There is an economic issue about whether lawyers with access to low-cost or free services such as Fastcase (often comes with bar subscription), Casetext (free plus premium service), Ravel (free for academics plus a premium service), and Google Scholar (free) can through iterative case searching achieve results truly comparable to Westlaw and Lexis. This needs to be studied.

44. Compare CALLISTER, FIELD GUIDE, supra note 28, at 19–24 with id. at 24–34. The example I use with known-item problems is “You need the California murder case in which the court found that a fetus cannot be a human being, and the defendant was acquitted of murder after beating up his wife or girlfriend resulting in the loss of the fetus.” Id. at 19, tbl. 3–2. This contrasts with a subject problem where “You are looking for an explanation of low-income housing credits.” Id. at 24, tbl. 3–3.

45. Usually, we have to filter to get to secondary sources.
the single-search box world. Later, this article will examine the issue and find that both Lexis Advance and Westlaw Edge have taken steps with predictive searching to preserve the role of secondary materials (although circumventing the use of indexes).  

¶19 It may give the librarian reader some comfort that not only is the behemoth that is Thomson Reuters (hereinafter Westlaw) sticking to its traditional strength with human "curated" information, but it even outperforms other systems when it employs AI—at least for the examples I provide in this article with respect to case law research. However, the issue is not just case law, where Westlaw has so much curated information or metadata, but the role of secondary sources and how we access them. And, we need to consider access to codes and whether natural language processing works for them. So along with indexes and human-intermediated information, are secondary sources in danger of extinction, at least in terms of use, if not cessation, in the near future? AI may one day replace the "secondary sources" with cognitive authority by writing its own summaries and expounding the law from primary sources. And if it can do it better than human experts, why not? Abstracting and summarizing is already being done with natural language processing. Even a book has been written by such a tool and published by Springer Nature. It is not a great stretch to apply to legal research and writing. Westlaw has 2.4 million state and federal briefs. This is big data that natural language processing can analyze to learn the structure and nuance of brief writing—if not now, soon. Brief analysis is also a feature of Westlaw Edge, Casetext, Ross Intelligence, and several other vendors’ products. Actually writing the brief may not be far behind. Indeed, as far back as 2005, Dan Dabney, Senior Director, Thomson Global Services GmbH, described a future in which attorneys would start writing

46. See infra ¶¶ 75–85.
48. See infra ¶¶ 53–74. This is also true for the Nevelow Mart study. See Nevelow Mart, supra note 34, at 414, chart 2 & tbl. 3 (Westlaw led in precision relevance results).
53. Similar predictions were raised by a New Yorker columnist about whether a machine could write for the New Yorker. See John Seabrook, The Next Word. Where Will Predictive Text Take Us?, NEW YORKER (Oct. 14, 2019), https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker (last visited Aug. 4, 2020). The technology spawning the article was the author’s interaction with predictive text using Smart Compose in Google Email. Id.
briefs, and the research would be seamlessly supplied to them. Quoting Dabney, “[W]hat is happening here, at least potentially, is that legal research has ceased to be a particularly separate part of the operation. You can just sit down and write a brief and the authorities you need, the law that you are looking for, will find you.” Natural language processing, potentially, could support such a future.

Building a Model of an Information Retrieval System with Natural Language Processing

The query is not the sole focus of natural language processing, although it is part of it. Increasingly, large databases are mined for their inner relationships among documents. This can be done independently of any search, with the purpose of finding relationships of words or documents to each other. Turning words into vectors is an important technique for this process. There are also other tools besides trigonometric vectors, involving probabilities and neural networks. Researchers experiment with all of these to use natural language processing to discover relationships among documents and words in databases. To understand a little more of this activity, we will start with understanding how word vectors (also known as embeddings) are employed.

Trigonometric Word Vectors and Cosine Similarity

The thing to understand is that once words have been turned into vectors (or dimensions), they can be compared, added or subtracted, and then, quite remarkably, the nearest word can be determined.

As a famous example, vectors for the following words can be processed on a linear basis:

king – man + woman = queen

What happens is the vectors as processed above result in a vector that is nearest the word queen. Essentially king is to man as queen is to woman. This is analogical computing, and it is done on what is known as a linear basis (meaning mathematical functions apply). All sorts of interesting relationships can develop with this or similar processes. In theory, vectors for purple – red = blue. Vectors for France – Paris + Athens = Greece. All of this is linear processing.

To understand the determination of vectors, I need to lay out some more concepts and two frequently used natural language processing models.

Suppose we have the following co-occurrences of terms:

---

55. Id., citing Dan Dabney, Envisioning the Future: The Publisher’s Perspective, Remarks at the Future of Law Libraries Symposium, Florida Coastal School of Law (Mar. 10–11, 2005) (quoted passage transcribed by author from digital recording no longer available over the Internet).
Table 1

<table>
<thead>
<tr>
<th></th>
<th>Rent</th>
<th>Sale</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warranty</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Lease</td>
<td>6</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Contract</td>
<td>0</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

We want to know whether warranty (including its stem, warrant) is closer to contract or lease (at least as word vectors). To do this, we need the cosines of our terms in relation to each other, or at least:

\[ \cos(\text{warranty}, \text{contract}) \text{ and } \cos(\text{warranty}, \text{lease}) \]

The formula for calculating the cosine of two vectors is:

\[
\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}
\]

In the numerator is the “dot product” of the two vectors, which is essentially multiplying two vectors (or each of their members in the set) by each other. In the denominator is the multiplication of each vector’s length times the other. I will illustrate later. Let’s start with the numerator. It can be calculated:

\[
\sum_{i=1}^{N} v_i w_i
\]

The denominator (or multiplication of vector lengths) is calculated by squaring each vector set member, adding them together, and then taking the square root. This is done for each set, the results of which are then multiplied by each other. The whole process is summarized:

\[
\sqrt{\sum_{i=1}^{N} v_i^2} \times \sqrt{\sum_{i=1}^{N} w_i^2}
\]

Consequently, with the data in the table above, we calculate the numerator for warranty and contract vectors:

\[
(3 \times 0) + (4 \times 6) + (3 \times 4) = 36
\]

And the denominator for warranty and contract vectors:

\[
\sqrt{3^2 + 4^2 + 3^2} \times \sqrt{0^2 + 6^2 + 4^2} = 42.04759
\]

The cosine requires that we simply divide the numerator and the denominator, which for warranty and contract equals 0.856173.
Repeating the step for the numerator of the warranty and lease vectors:

\[(3 \times 6) + (4 \times 0) + (3 \times 2) = 24\]

And the denominator for warranty and lease vectors:

\[\sqrt{3^2 + 4^2 + 3^2} \times \sqrt{6^2 + 0^2 + 2^2} = 36.87818\]

Again, the cosine requires that we simply divide the numerator and the denominator, which for warranty and lease equals 0.650791. Thus comparing the two cosines, warranty has greater similarity to contract than it does to lease (at least within our hypothetical database with its small set of data).

¶25 These vectors involve combining vectors for single words, with at least three dimensions (rent, sale, and credit). Extracting cosines can help us cluster words and identify topics in documents or sentences. In large search databases, like Google or Federal Supplement on Lexis or Westlaw, the dimensions may be much larger.

¶26 As we scale up for larger databases:

It turns out, however, that simple frequency isn’t the best measure of association between words. One problem is that raw frequency is very skewed and not very discriminative. If we want to know what contexts are shared by [the list of terms], we’re not going to get good discrimination from words like the, it, or they, which occur frequently with all sorts of words and aren’t informative about any particular word. . . . the [dimension] for the word good is not very discriminative between [Shakespeare’s] plays; good is simply a frequent word and has roughly equivalent high frequencies in each of the plays.59

The problem addressed above can be addressed by methods I will introduce below. It can also be addressed by filtering stop words.60 The problem with stop words is identifying them ahead of time. In law, there are lots of terms whose frequency across documents in a database may not be helpful in classifying documents—for example, law, legal, court, judge, ruling. We can either create a list of stop words or decrease the weight of such terms. The latter is preferable because stop words have an absolute rather than graduated effect. The next section will treat this topic by showing a more sophisticated technique.

**TF-IDF or Term Frequency and Inverse Document Frequency**

¶27 The basic idea behind stop words is to give less weight to terms that occur too frequently in a document, especially if they also appear frequently across all of the documents of a database. Logarithmic scales (based on orders of magnitude) are used to give weights to words and inverse weights if words appear frequently in documents across the database. For the first part of the calculation, a word that appears 10 times in a document would have a weighted term frequency (tf) of 2, and one that appeared 100 times would have a weighted tf of 3. These tfs are then multiplied by the inverse of the term in document frequency (also on an logarithmic scale), known as idf.

59. See id. at 12.

¶28 Here is the formula for term weight:\(^{61}\)

\[ w_{t,d} = tf_{t,d} \times idf_t \]

So essentially the weight of the term in a given instance of a document is the product of the term frequency in the document (which will be defined below) and the inverse of the term in all documents (which is also defined below). To calculate the term frequency in a document, we use a logarithmic function:

\[ tf_{t,d} = \begin{cases} 1 + \log_{10}\text{count}(t,d) & \text{if } \text{count}(t,d) > 0 \\ 0 & \text{otherwise} \end{cases} \]

So term frequency in a document is determined by first deciding whether the count of the term in the document is greater than zero. If so, the term frequency is represented by adding one to the base 10 logarithm of count of the term in the document. So if the term appears 5 times in the document, \(1 + \text{the base 10 logarithm of 5}\) equals 1.698970004.

¶29 The next step and factor is designed “to give a higher weight to words that only occur in a few documents. Terms that are limited to a few documents are useful for discriminating those documents from the rest of the collection.”\(^{62}\) To calculate the inverse document frequency (\(idf\)) for a term:

\[ idf_t = \log_{10} \left( \frac{N}{df_t} \right) \]

\(N\) is the number of documents in the database, and \(df_t\) is the frequency of documents with the term. Thus the number of documents in the database is divided by the frequency of documents with the term. Then a logarithm base 10 is determined. So if the number of documents in the database is 10, and the number of documents with the term is 5, the result is the logarithm base 10 of 2 \((10/5)\), which is 0.301029996.

¶30 Ultimately, we must determine the product of the previous two functions to give us the appropriate weight, or in our example 0.511440933.

¶31 We might use the following hypothetical data to illustrate term frequencies across different documents in a database of 1000 documents.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>law</td>
</tr>
<tr>
<td>legal</td>
</tr>
<tr>
<td>tax</td>
</tr>
<tr>
<td>credit</td>
</tr>
<tr>
<td>housing</td>
</tr>
<tr>
<td>allocate</td>
</tr>
</tbody>
</table>

61. For math that follows, see JURAFSKY & MARTIN, supra note 58, at 13–14.
62. Id. at 13.
Using an Excel spreadsheet, the appropriate weight for the terms in Document 1 can be calculated as follows:

Table 3

<table>
<thead>
<tr>
<th>Term</th>
<th>Term Frequency in Document 1</th>
<th>Document Frequency for Term</th>
<th>1+Log10 for tf</th>
<th>log10(n/df) or idf</th>
<th>Number of Docs in Database (N)</th>
<th>Product (wtd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>law</td>
<td>30</td>
<td>1000</td>
<td>2.48</td>
<td>-</td>
<td>1,000</td>
<td>-</td>
</tr>
<tr>
<td>legal</td>
<td>5</td>
<td>900</td>
<td>1.70</td>
<td>0.05</td>
<td>1,000</td>
<td>0.08</td>
</tr>
<tr>
<td>tax</td>
<td>105</td>
<td>90</td>
<td>3.02</td>
<td>1.05</td>
<td>1,000</td>
<td>3.16</td>
</tr>
<tr>
<td>credit</td>
<td>12</td>
<td>130</td>
<td>2.08</td>
<td>0.89</td>
<td>1,000</td>
<td>1.84</td>
</tr>
<tr>
<td>housing</td>
<td>12</td>
<td>5</td>
<td>2.08</td>
<td>2.30</td>
<td>1,000</td>
<td>4.78</td>
</tr>
<tr>
<td>allocate</td>
<td>7</td>
<td>85</td>
<td>1.85</td>
<td>1.07</td>
<td>1,000</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Here we have calculated the appropriate weight of a term in a document (wtd) using inverse factors for high frequency of documents with a term across the database (idf). Logarithmic functions have facilitated this. The size of our database, or N, is 1000 documents.

\[ \text{¶}32 \text{ We repeat this process for all four documents and get the following weights across the four documents:} \]

Table 4

<table>
<thead>
<tr>
<th>Term Weights tf-idf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
</tr>
<tr>
<td>law</td>
</tr>
<tr>
<td>legal</td>
</tr>
<tr>
<td>tax</td>
</tr>
<tr>
<td>credit</td>
</tr>
<tr>
<td>housing</td>
</tr>
<tr>
<td>allocate</td>
</tr>
</tbody>
</table>

Notice that law has no weight. This is because it was found in every document in the database (all 1000). As we might expect, legal has very low weight because of its frequency (900/1000).

\[ \text{¶}33 \text{ From these term weights, we can now rank which documents are most similar and dissimilar based on the terms (which is very useful in natural language searching). We can do this by determining the comparative cosines of each of the documents based on their shared terms. For Documents 1 and 2, we make a } \text{dot product} \text{ calculation:} \]

\[ (0.08 \times 0.09) + (3.16 \times 1.99) + (1.84 \times 2.35) + (4.78 \times 5.34) + (1.98 \times 2.44) = 41.01 \]

\[ \text{¶}34 \text{ We then calculate the square root of the sum of each term frequency squared. We do this for each document. For Document 1,} \]

\[ \sqrt{0.08^2 + 3.16^2 + 1.84^2 + 4.78^2 + 1.98^2} = 6.34 \]
For Document 2, 

\[ \sqrt{0.09^2 + 1.99^2 + 2.35^2 + 5.34^2 + 2.44^2} = 6.63 \]

We then perform the following operation:

\[ \frac{41.01}{6.34 \times 6.63} \]

The result is a combined cosine for Documents 1 and 2 of 0.98. This means that with respect to the terms we have identified, the documents are almost identical. We can make those comparisons with all of the documents.

Table 5

<table>
<thead>
<tr>
<th>Summary of Cosines</th>
<th>Doc 1</th>
<th>Doc 2</th>
<th>Doc 3</th>
<th>Doc 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>1.00</td>
<td>0.98</td>
<td>0.65</td>
<td>0.13</td>
</tr>
<tr>
<td>Doc 2</td>
<td>0.98</td>
<td>1.00</td>
<td>0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Doc 3</td>
<td>0.65</td>
<td>0.53</td>
<td>1.00</td>
<td>0.64</td>
</tr>
<tr>
<td>Doc 4</td>
<td>0.13</td>
<td>0.16</td>
<td>0.64</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Based on this, we can see that Document 2 is most like Document 1. Document 3 is the next nearest to Document 1, and Document 4 is not very similar at all to Document 1. We haven’t exactly created a search engine, but we have a model that would be useful in creating one. Uses of vector comparisons between documents might include Westlaw folder analysis, where based on the documents in a folder Westlaw Edge suggests additional similar documents.63 Also, it could be used to compare cases for matching with Westlaw Edge’s headnotes and Topic and Key Number System, and for Lexis’s topic classification system (including uses with “More like this Headnote”). In addition, Westlaw Edge’s new overruling risk could work by locating documents with similar cosines to a document that has been expressly overruled.64 I am sure there is more to it than is outlined here, but vector cosines are a powerful tool in helping us think about how our major database services work.

Centroids and Document Similarity

¶35 There is a simple technique for ranking the documents based on the term weights (rather than cosines), and that is to take the average of the sum of term weights for each document. This figure is known as the centroid, and can be used to rank documents in order.

63. See Thomson Reuters, Not All Legal AI Is Created Equal, supra note 47.
Table 6

<table>
<thead>
<tr>
<th>Term</th>
<th>Doc 1</th>
<th>Doc 2</th>
<th>Doc 3</th>
<th>Doc 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>law</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>legal</td>
<td>0.08</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>tax</td>
<td>3.16</td>
<td>1.99</td>
<td>2.77</td>
<td>1.54</td>
</tr>
<tr>
<td>credit</td>
<td>1.84</td>
<td>2.35</td>
<td>1.51</td>
<td>1.93</td>
</tr>
<tr>
<td>housing</td>
<td>4.78</td>
<td>5.34</td>
<td>–</td>
<td>2.99</td>
</tr>
<tr>
<td>allocate</td>
<td>1.98</td>
<td>2.44</td>
<td>1.07</td>
<td>1.82</td>
</tr>
<tr>
<td>centroid</td>
<td>2.37</td>
<td>2.44</td>
<td>1.09</td>
<td>1.68</td>
</tr>
</tbody>
</table>

Looking at the centroids, we can see that Document 2 is closest to our terms, and Document 1 is a close second. Document 4 is a distant third. The Summary of Cosines is helpful in a search because once a document has been identified as relevant, it is easy to discover what other documents might be relevant.

§36 Perhaps the most common way to use our packet of tools is to calculate the cosines of words across documents (we have already calculated the cosines of documents across words). Imagine doing this across all 1000 documents in our hypothetical database (or across all the case law in Lexis’s Federal Supplement database) and using a vocabulary that extends into the tens or hundreds of thousands of words.65 We can, quite handily, find the synonyms for legal terms by finding the closest cosines using this technique.66 This can be helpful in expanding the initial query to appropriately related concepts and topics.

Queries as Vectors—An Example

§37 Not only can documents and words be vectors, but queries can be vectors.67 Imagine the calculations of the following query for tax, credit, housing, and allocate.68 We can calculate the TF-IDF term weights as follows:

Table 7

<table>
<thead>
<tr>
<th>Term</th>
<th>Term Frequency in Query 1</th>
<th>Document Frequency for Term</th>
<th>$1 + \log_{10}$</th>
<th>$\log_{10}(n/dft)$ or idft</th>
<th>Number of Docs in Database (N)</th>
<th>Product (wtd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>law</td>
<td>0</td>
<td>1000</td>
<td>0</td>
<td>–</td>
<td>1,000</td>
<td>–</td>
</tr>
<tr>
<td>legal</td>
<td>0</td>
<td>900</td>
<td>0</td>
<td>0.05</td>
<td>1,000</td>
<td>–</td>
</tr>
<tr>
<td>tax</td>
<td>1</td>
<td>90</td>
<td>1</td>
<td>1.05</td>
<td>1,000</td>
<td>1.05</td>
</tr>
<tr>
<td>credit</td>
<td>1</td>
<td>130</td>
<td>1</td>
<td>0.89</td>
<td>1,000</td>
<td>0.89</td>
</tr>
<tr>
<td>housing</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>2.30</td>
<td>1,000</td>
<td>2.30</td>
</tr>
<tr>
<td>allocate</td>
<td>1</td>
<td>85</td>
<td>1</td>
<td>1.07</td>
<td>1,000</td>
<td>1.07</td>
</tr>
</tbody>
</table>

65. See Jurafsky & Martin, supra note 58, at 15.
66. Id. (“[W]e can find the 10 most similar words to any target word $w$ by computing the cosines between $w$ and each of the $V$-1 other words, sorting, and looking at the top 10.”).
68. In a real search engine, word stems for the terms might be used.
We now have a table of term weights that looks like the following:

<table>
<thead>
<tr>
<th>Term</th>
<th>Doc 1</th>
<th>Doc 2</th>
<th>Doc 3</th>
<th>Doc 4</th>
<th>Query 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>law</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>–</td>
<td>0.00</td>
</tr>
<tr>
<td>legal</td>
<td>0.08</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>tax</td>
<td>3.16</td>
<td>1.99</td>
<td>2.77</td>
<td>1.54</td>
<td>1.05</td>
</tr>
<tr>
<td>credit</td>
<td>1.84</td>
<td>2.35</td>
<td>1.51</td>
<td>1.93</td>
<td>0.89</td>
</tr>
<tr>
<td>housing</td>
<td>4.78</td>
<td>5.34</td>
<td>–</td>
<td>2.99</td>
<td>2.30</td>
</tr>
<tr>
<td>allocate</td>
<td>1.98</td>
<td>2.44</td>
<td>1.07</td>
<td>1.82</td>
<td>1.07</td>
</tr>
</tbody>
</table>

From there we can determine combined cosines for Query 1 and each of the documents using the techniques described above.

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>Doc 2</th>
<th>Doc 3</th>
<th>Doc 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query1</td>
<td>0.9879</td>
<td>0.9968</td>
<td>0.5594</td>
</tr>
</tbody>
</table>

The ranking of documents in a natural language search based on our query is thus: Document 2, Document 1, Document 4, and Document 3, with Documents 2, 1, and 4 being near perfect vector matches.

¶38 Note that the query terms in Table 7 are represented by a 1 or a 0. The zero means we are not including a term from the vocabulary of terms in the database. There would be thousands or tens of thousands of zeros for any given search. Also note that natural language search terms rarely include the term twice, but if a search phrase repeated a term, it could affect the weight of the term in the search, but on a logarithmic scale, and the ultimate ranking of results. This gives the user just a little control in emphasizing certain terms in search results by repeating terms.

¶39 I have presented only the simplest versions of plausible natural language processing with our major online services. Yet, cosine calculations using TF-IDF with database vocabularies of thousands of words (or dimensions if we think in terms of vectors) are voluminous calculations that are costly for computing time. Personally, I marvel to think what might go into our legal search engines.

---

69. See tbl. 7 (the column for $1+\log_{10}\text{for } tf$ would be affected on a logarithmic scale).
70. For an additional technique to add weight, see infra note 139 and accompanying text.
71. Manning, Raghavan & Schütze, supra note 67, at 114.

Computing the cosine similarities between the query vector and each document vector in the collection, sorting the resulting scores and selecting the top $K$ [ranked] documents can be expensive – a single similarity computation can entail a dot product in tens of thousands of dimensions, demanding tens of thousands of arithmetic operations.

Id.
¶40 There is a limitation with this kind of model. It assumes that co-occurrence of terms in a legal document establishes a relationship that is relevant to the search query. Obviously, if we consider case law, many issues (which may be unrelated) may be found in a single case. Think of cases that are unrelated for 95% of the content, but on the precise issue in question they are related. Now, if the co-occurrence were limited by context (i.e., a range of words proximate to a targeted term), rather than co-occurrence scattered throughout the whole document, we might get more relevant search results. Our next model addresses this issue.

word2vec and Skip-gram with Negative Sampling

¶41 Cosine calculations using TF-IDF use thousands and tens of thousands of references (called dimensions) of co-occurring words in a document. The resulting vectors or embeddings are said to be “sparse,” because most of the weights are zero. However, using probabilities instead of cosines employs far fewer dimensions to be used with each word, and are said to be “dense,” with many more positive weights. As few as 50 dimensions can be used, but the ranges often extend into the hundreds (rather than the tens of thousands).72 This simplifies calculations and produces more accurate results.

¶42 Skip-gram with negative sampling (SGNS) is one of two techniques for dense vectors offered by a research package known as word2vec.73 “Skip-gram predicts surrounding words given the current word.”74 It works by introducing words “into the vector one at a time, and scanning back and forth within a certain range.”75 Context of terms is important to the technique.

¶43 Because of the lengthy complexity of the math (and, candidly, my inability to understand it in other than the most general terms), I will not illustrate it here but, essentially,76 a target word is identified with a given number of words to the

---

72. See Jurafsky & Martin, supra note 58, at 18. This is relatively few compared to the embeddings or dimensions described in note 66, supra, and accompanying text.
74. Mikolov et al., supra note 56, at 5, fig.1. In essence, a skip-gram is the opposite of another tool offered by word2vec known as CBOW (Continuous Bag of Words), which predicts the target word based on the context. Id. The skip-gram method predicts the context words and “produces more accurate results on large datasets” than CBOW. Skymind, A Beginner’s Guide to Word2Vec and Neural Word Embeddings, A.I. Wiki, https://skymind.ai/wiki/word2vec (last visited Aug. 4, 2020). The entire text corpus is processed. Id.
75. Skymind, supra note 74.
76. See Jurafsky & Martin, supra note 58, at 19–23. Dot products for cosines are turned into probability using a sigmoid function:

$$
\sigma(x) = \frac{1}{1 + e^{-x}}
$$

Id. at 20, formula 6.25. The function helps determine the probability for the context word for a given target word as well as for random words not being context words (\(-x = -t\cdot c\)). See also Xin Rong, word2vec Parameter Learning Explained (June 5, 2016), https://arxiv.org/pdf/1411.2738 (last visited Aug. 4, 2020).
right and left of it, known as context words. Probabilities are calculated for context words. These are considered positive examples of dimensions or embeddings. An even larger set of negative examples is generated at random from the lexicon for the database. These words are known as “noise words.” Using probabilistic statistics and statistical regression (and assuming the probabilities for positive and negative examples equals one), weights for embeddings are created for the words in the document. Not only does SGNS yield shorter embeddings, but it does a better job of generalizing and capturing synonyms.77

¶44 SGNS works according to the following general rules:

1. Treat the target word and neighboring context word as positive examples;
2. Randomly sample other words in the lexicon to get negative samples;
3. Use logistic regression to train a classifier to distinguish those two cases; and
4. Use the regression weights as embeddings.78

Classifiers will be treated below,79 but basically the end results are weights for positive and negative examples of the probabilities, which collectively equal one. Having produced embeddings for document databases, skip-gram tools can compare the results for similarity of embeddings for search phrases.

Potential Use of word2vec in Information Retrieval

¶45 As with cosine vectors, the potential exists for vectors coming from word2vec to be used with information retrieval. However, such vectors have been primarily used in word clustering or topic modeling.80 But there are exceptions.

For instance, if we obtain vector representations of a collection of texts we can apply clustering algorithms directly to these representations to automatically group similar documents together to facilitate searching through a large corpus. Or we can apply supervised learning models that predict an outcome [such as passage of a bill] to the text. The possibilities are almost endless.81

¶46 Other data scientists have directly proposed that word2vec vectors or embeddings be used in information retrieval.82 Unfortunately, we simply do not know whether such tools are used with the major legal databases but, as shall be seen below,83 we can make some educated guesses.

---

77. See Jurafsky & Martin, supra note 58, at 18.
78. Id. at 19.
79. See infra ¶¶ 48–50.
83. See infra ¶¶ 53–74.
Regression and Neural Networks

§47 One of the most advanced techniques for determining embeddings is the use of neural networks (see figure 2). Such computer networks test regressions of variables for terms until the optimum is reached. The number of layers adds to the complexity and computing power of the network. Describing their function other than for use in regression and classification is beyond the scope of this article. The nodes do tend to work by recognizing patterns and using inference, rather than instruction. It is unknown whether the major vendors use neural networks.

Classification

§48 Besides word vectors, some of the tools of natural language processing come from statistics and include Bayes classification. A basic use is to determine whether text or a document falls within a class. For example, is an email spam? Another use is to determine sentiment—is a movie review positive? In our context of legal research, we can imagine that the major vendors—Lexis, Westlaw, and Bloomberg—could use such techniques to know whether a case has negative treat-

---


ment or is distinguished in another case, or with classification in Westlaw Edge’s Key Number System, or Lexis Advance’s “More like this Headnote.” Of course, because these systems are proprietary, we do not know for certain which natural language processing tools are being used.\footnote{See supra note 8.} Nonetheless, it is enlightening to study how various natural language processing tools work.

Suppose we are trying to train a case law database natural language processing tool to assess whether a case has negative treatment. We might use a training database of expression that looks like the following:

<table>
<thead>
<tr>
<th>Training Category</th>
<th>Document Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>clearly distinguished by</td>
</tr>
<tr>
<td>Pos</td>
<td>this case was effectively overturned</td>
</tr>
<tr>
<td>Pos</td>
<td>the facts do not set on all fours with</td>
</tr>
<tr>
<td>Neg</td>
<td>closely followed by</td>
</tr>
<tr>
<td>Neg</td>
<td>confirms the holding from</td>
</tr>
<tr>
<td>Test</td>
<td>case distinguished from</td>
</tr>
</tbody>
</table>

Essentially, phrases in the positive category indicate that subsequent treatment in a case is negative treatment of the original cases. We are going to apply an algorithm known as Naive Bayes Classifier to test the bottom document phrase in the table. We first take the likelihood of negative and positive categories in our dataset of sentence phrases.\footnote{See Jura\,fsky \& Martin, supra note 86, at 7.}

\[
P(+) = \frac{2}{5} = 0.4
\]

\[
P(\neg) = \frac{3}{5} = 0.6
\]

We next need the likelihood of the three words in the test phrase being classed as positive or negative.\footnote{Id.}

\[
P(\text{"case"}|+) = \frac{\text{count}(w,c) + 1}{\sum_{w \in V} \text{count}(w,c) + |V|} = \frac{1 + 1}{8 + 17} = 0.08
\]
Here, \( w \) is the word, \( c \) is the class illustration of positive or negative phrases, and \( |V| \) is the union of all word types in all classes (i.e., no counting of duplicate words across the class). The two classes are then multiplied by each result for the class (i.e., all positive classes multiplied against each other, and all negative classes against each other). Finally, the results (the probability of our test sentence phrase) are multiplied by the likelihood of the positive and negative categories.\(^91\)

\[
P(+)P(S+) = 0.04 \times (0.08 \times 0.08 \times 0.04) = 1.024 \times 10^{-4}
\]

\[
P(−)P(S−) = 0.06 \times (0.037037 \times 0.037037 \times 0.074074) = 6.0966 \times 10^{-5}
\]

\(^50\) It is more probable that our test sentence phrase (\( S \)) represents a positive instance of negative case history. Thus, we can see the mechanics of how classification of negative case history occurs. It is all based on probability comparisons. This is a far departure from the human expert system established by Shepard’s, but coded classifications of when a case has negative treatment is doable as KeyCite and as the above exercise in the probabilities of a classification system have each demonstrated. Furthermore, such systems depend on some human training. In a blog post on *Above the Law*, David Lat confirmed that Thomson Reuters’ KeyCite Overruling Risk (which tracks not only the original cases, but related cases whose overruling may affect the original case) relies upon “natural language processing as well as traditional, supervised machine learning [probably referring to human training and testing].”\(^92\) Well, we do not know for certain which natural language processes were

91. *Id.*
92. David Lat, *How Artificial Intelligence Is Transforming Legal Research, Above the Law* 2020,
used; however, the Bayesian exercise we have gone through, or related statistical tools, could be adapted for use. It is important to understand that humans train such systems, but then the systems are turned loose for disintermediated service to the legal profession. How much do we trust opaque natural language processes to approximate a legal mind in identifying negative precedent, or is that even the right question? Perhaps the real issue is for data scientists who can compare results from natural language processing systems against one another.

Grammatical Techniques

¶ 51 Another use of Bayesian techniques is to combine them with Hidden Markov Models to identify parts of speech and phrases of particular value. 93 One paper on the topic connected sentence structure to “better extraction accuracy,” 94 referring to information extraction of “specified classes or relations from text.” 95 With such techniques, sentence structure, rather than co-occurrence of related terms, matters.

¶ 52 Nevelow Mart has distinguished between key word searching (like when two or three terms are entered into Google) and natural language searching (which is aided by “grammatical structure models”). 96 I am not exactly sure what is meant by “grammatical structure models,” but statistical techniques employing Hidden Markov Models have been employed to match words with their “parts of speech.” 97 For some time, such techniques have also been used for “information extraction” including “summarizing collections of documents, and identifying significant but unknown relations among objects” and for information retrieval systems. 98 Besides parts of speech, other grammatical rules may be in play. One article from 2003 on the subject noted that “grammatical rules contribute to most of the behavior of Natural Language parsers.” 99 However, parsing long sentences, even using a complex set of grammatical rules, is avoided due to difficulty. 100 That said, generally, employment of grammatical rules and Hidden Markov Models are beyond the scope of this article, which instead focuses on word vectors (and given the results


94. Ray & Craven, supra note 93, at 2.
95. Id. at 1.
96. See Nevelow Mart, supra note 34, at 397, ¶ 16.
98. See Ray & Craven, supra note 93, at 1.
100. See id.
of the next section, which breaks up sentence structure into test searches, this emphasis turns out to be appropriate).

**Bag of Words, word2vec, and the Performance of Current Legal Search Engines**

§53 The techniques of TF-IDF vector comparisons and word2vec skip-grams with negative sampling (SGNS) that we have been using in §§41–44 treat all terms as a “bag of words,” meaning word order and proximity (at least within the range of context words for SGNS) do not matter. We can ask the question whether this is the case in a review of major and minor case law legal research services. In investigating the issue of “bag of words,” I have done some unusual things—run a search on a phrase backwards and then scramble it to see whether similar results would appear. I do not claim this to be the most scientific and disciplined study (it grew organically from a single problem); rather, the activity illustrates, using a fairly difficult and real research question, some of the issues we presently face and the extent to which vendors may incorporate natural language processing. I will explain the research issue more fully as we go through the results.

**Westlaw Edge**

§54 Table 11 illustrates a fairly sophisticated search, where we then reverse the order of the search terms and finally scramble them. We then compare the top 10 results (at least for precision)\(^{101}\) to see whether Westlaw Edge employs a “bag of words” vector approach.\(^{102}\)

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c) as a health and disability plan</td>
<td>Beisler v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Rosen v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gibson v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Craft v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Abbate v. Spear</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Estate of Barnhorst v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Beisler v. Comm’r (first hearing)</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Hines v. Comm’r</td>
<td>N</td>
</tr>
</tbody>
</table>


102. Except as noted, the date of the following searches was July 9, 2019. I have already discovered that rerunning searches at later dates provides slightly different results. This raises interesting questions about what is changing—perhaps the systems are reacting to training in the form of aggregating user responses. Nevelow Mart documents that Fastcase “aggregates history of more than 100 million searches.” Nevelow Mart et al, *supra* note 8, at 13. Probably all of the services take advantage of aggregated user history.
The issue I am researching lies at the intersection of pension law and two differing tax code schemes for plans that are excluded from income. One excludes retirement plan contributions and earnings; the other excludes health/accident plan contributions, earnings, and distributions based on the nature of the injury. As a former tax and pension attorney, I can state that the goal of some tax attorneys is to combine the two different kinds of plans (or at least claim that a retirement plan also operated as a health/accident plan). It is a tough issue to research and requires vigorous inquiry into case law.

¶55 Beisler and Caplin were two of the cases I relied upon heavily in practice and in publication on this topic, although there were many more cases that addressed the issue. Note the results are very similar regardless of how the query is entered—even if backwards or scrambled. In fact, the backwards and scrambled queries have only one irrelevant result each (marginally better than the straightforward search). This stresses how little word order and proximity (within the context range for SGNS) may matter to the search engine. This suggests use of word vectors (whether determined as cosines of co-occurrence or as probabilities based on context). Also note, however, the lists, while very close, are not exactly identical. An

---

The research director at Thomson Reuters freely admits that tinkering with the algorithm may produce different results in terms of recall and precision—some algorithms favoring the former and some the latter:

As Curtis explains, “In technological terms, it’s the trade-off between precision—finding the right things—versus recall—finding all the things. If we tune an algorithm to score high in precision by delivering the ‘best’ results, there’s a trade-off in terms of some relevant results getting missed—in other words, lower recall. On the other hand, if we tune an algorithm to score high in recall by not missing relevant results, the best results will be somewhere in there, but not at the top, or maybe even buried—in other words, lower precision.”

This balancing act affects two very different measures of search success—precision measures how many results from an inquiry are relevant, but recall measures how many relevant documents were found compared to relevant documents in the database as a whole. Furthermore, what Curtis is acknowledging is that you cannot have it both ways—both high precision and recall scores.

Additionally, the range of the context window for a targeted term in a skip-gram is subject to adjustments for words that are too frequent or too few in the database. Thus, the window size is dynamic. Indeed, the size of the context window (known as the $k$ variable) is set by the algorithm’s designer. How big a context window is can have a huge impact on results:

Shorter context windows tend to lead to representations that are a bit more syntactic, since the information is coming from immediately nearby words. When vectors are computed from short context windows, the most similar words to a target word $w$ tend to be semantically similar words with the same parts of speech. When vectors are computed from long context windows, the highest cosine words to a target word $w$ tend to be words that are topically related but not similar.

There are many variables that have to be selected for a search algorithm—it is a balancing act—and they may make a tremendous difference in terms of results. Thus, we should expect a significant variation in results among legal research services.

Lexis Advance

The results in Lexis Advance suggest that it is probably not using a “bag of words” vector comparison approach and that word order and, perhaps, proximity matter. The initial results are less than stellar for this complex search. In fact, only
eight cases were identified by the first search, of which four were relevant (and none of these were Beisler and Caplin, two cases that Westlaw ranked most prominently, and which I heavily relied upon in practice and in publication). 112 Furthermore, reversing the terms in the search query results in totally different results with no relevant cases. 113 Based on the initial search in July 2019, the order of search terms seems to matter in a Lexis Advance search, suggesting that the “bag of words” word vector approach may not be in use. However, a subsequent search, in November 2019, produced very different results: 114 more relevant results and key cases were recorded on the initial search; reversing the search terms still produced no results; and finally, scrambling the search terms yielded a few relevant cases, although very different search results from the straightforward search. The conclusion is that word order still matters in Lexis Advance searching, 115 and there is less, if any, reliance on a “bag of words” vector approach. There also appears to be little stability in search results. I have found that even running a search a few days later gives new results (perhaps because the system is reacting to my behavior as a user). 116

¶ 60 That reversing the order of the query causes a return of entirely irrelevant results is convincing evidence that “bag of words” and word vectors are not being used, but proximity and word order are important.

Table 12

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c) as a health and disability plan</td>
<td>ABA Ret. Funds v. U.S.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Antioch Co. Litig Trust v. Morgan</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Kifafi v. Hilton Hotels Ret. Plan</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Fr. v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Wright v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Chernik v. Comm’r</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Minnequa Univ. Club v. Comm’r</td>
<td>N</td>
</tr>
</tbody>
</table>

112. See supra note 103.
113. At the time of the original search, because of the lack of relevant results, there seemed to be no point to scrambling the search terms as I did with Westlaw Edge.
114. See appendix A. The initial search was done on July 9, 2019. Since the initial search, more relevant results appear in the straightforward search query. See appendix A (searched Nov. 13, 2019). Indeed, every time I run the same “scrambled” search, the results change and the precision often improves. Compare appendix A and appendix B (searched Nov. 18, 2019).
115. Indeed, techniques like use of Bayesian methods with Hidden Markov Model require ordered inputs or observations to predict outputs or “emissions.” Such techniques predict a probable path of emission such as the structure and order of a sentence. See discussion of Hidden Markov Models, supra ¶¶ 51–52. Of course, it is pure speculation (without any evidence) whether natural language processing in Lexis Advance might employ such methods.
116. Compare appendix A and appendix B.
¶61 To get Lexis Advance to improve its relevancy ratings, I had to revert to a terms and connectors search (outside the realm of natural language processing). The search terms were:

(“profit-sharing plan” OR “401(k) plan” OR “pension plan”) AND 105(c)

What I have done is replace “qualified retirement plan” (something Westlaw Edge was able to parse, probably by comparing vectors of terms to find terms that “qualified retirement plan” included) with specific types of retirement plans. I also had to delete references to “health and disability plans.”

¶62 Interestingly, in terms of precision,117 the results of Lexis Advance’s terms and connectors search were on par with Westlaw Edge’s natural language results (9 out of 10 results are relevant, although Beisler was missing). However, Lexis Advance’s results through terms and connectors are outside the inquiry of this article, which examines the scope of natural language processing and AI. What is really noteworthy, in comparing Lexis Advance and Westlaw Edge, is the ability of West to identify “profit-sharing plan” and other types of retirement plans with “qualified retirement plans.”

Table 13

<table>
<thead>
<tr>
<th>Search Query (terms and connectors)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>(“profit-sharing plan” OR “401(k) plan” OR “pension plan”) AND 105(c)</td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gibson v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Hall v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Zardo v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Enloe v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Rosen v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Cardinale v. S. Homes of Polk Cty.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Christensen v. U.S.</td>
<td>Y</td>
</tr>
</tbody>
</table>

117. Results in the first 10 hits that were relevant.
plan”—the more abstract term for all such plans in general. This may be the result of vector or embedding comparisons among terms.

**Bloomberg Law**

§63 Bloomberg Law also has some rather notable results.

**Table 14**

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c)</td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td>as a health and disability plan</td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Wood v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Green v. Comm’r</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Castellano v. City of N.Y.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Armstrong v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Atkins v. Bert Bell/Pete Rozelle NFL Ret. Plan</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Watts v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Hannington v. Sun Life &amp; Health Ins. Co.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Lovely-Beyea v. Me. State Ret. Sys.</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (backwards)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>plan disability and health a as 105(c) meeting also plan retirement qualified</td>
<td>Grose v. Grose</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Wood v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Stewart v. U.S.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Penn Allegh Coal Co. v. Holland</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Saunders v. Teamsters Local</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Castellano v. City of N.Y.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Marrs v. Motorola Inc.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Atkins v. Bert Bell/Pete Rozelle NFL Player Ret. Plan</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (scrambled)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>and retirement qualified plan also a 105(c) meeting disability as health</td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Green v. Comm’r</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Wood v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Estate of Hall v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Barnhorst v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Christensen v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Maller v. Comm’r</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
</tbody>
</table>
For Bloomberg, the best-performing group is when we scramble the search query (8 of 10 results are relevant, including the seminal case, Caplin). In comparison, the first search with the query in a straightforward order yields only 5 of 10 as relevant. Reversing the search algorithm gets only 3 of 10 relevant. What is going on here? Only the data scientists at Bloomberg know for sure, but the fact that we can scramble our search terms into a phrase with no syntactical cohesion and still get good (even better) results suggest there must be a role for word vectors and “bag of words.”

Fastcase

§64 Fastcase’s search engine cannot parse 105(c), and so the search term was shortened to 105, which certainly results in more “noise.” It also could not produce any relevant hits on the initial, straightforward query.

Table 15

<table>
<thead>
<tr>
<th>Search Query (federal appellate cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105 as a health and disability plan</td>
<td>Ashcroft v. Iqbal</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Bell Atl. Corp. v. Twombly</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Erikson v. Pardus</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Johnson v. U.S.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Lucia v. SEC</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Harrington v. Richter</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Sessions v. Dimaya</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Pearson v. Callahan</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Pearson v. Callahan120</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Molina v. Astrue</td>
<td>N</td>
</tr>
</tbody>
</table>

§65 We may want to stop here since we didn’t get any relevant results, but it might be interesting to determine whether we get the same irrelevant results by reversing the search terms or scrambling. Notably, the reversing of the search terms or the scrambling of them (omitting starting with and) produced identical results as the initial query, except the order of results is slightly different for the last three results.

118. The context width, or $k$ value, may be quite large since, as suggested above, large context ranges result in topical relatedness rather than semantic similarity. See supra note 110 and accompanying text.

119. The test for Fastcase was run on Nov. 7, 2019, a different date than the other databases. This is because during the initial test on July 9, 2019, I did not notice that natural language searching is not the default if entering terms into the “Quick Caselaw Search” bar.

While the reader may question the efficacy of disclosing search results with no relevant hits, the point is that the nearly identical results indicate that Fastcase may be using a “bag of words” vector approach to its natural language processing. Word order and proximity do not matter. However, none of Fastcase’s results were tax cases, which makes Westlaw Edge’s results even more remarkable. Westlaw’s algorithm detected, even though there was no reference to tax law or title 26 of the U.S. Code, that the relevant cases were all tax and pension cases through a natural language search. Although not producing as many relevant hits, Bloomberg Law picked up some tax cases. Lexis also located some tax cases, although only on a straightforward natural language query and a terms and connectors query (outside the scope of natural language processing).

121. "According to Fastcase’s promotional material, ‘natural language searches are much less precise’ than Boolean searches, ‘but they are a good place to start if you are new to legal research, or if you are delving into a new area of the law.’” Nevelow Mart, supra note 34, at 402, ¶ 24.
**Casetext**

%67 Casetext produces some interesting results in its straightforward query.\(^{122}\)

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c) as a health and disability plan</td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Estate of Barnhorst v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>In re Eagle Food Ctrs.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>In re Diet Drugs</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Beisler v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Crouch v. Brase Elec. Contracting Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Rashiel Salem Enters. v. Bunton</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>In re Phillips</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Chevron Corp. v. Barrett</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (backwards)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>plan disability and health as 105(c) meeting also plan retirement qualified</td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Estate of Barnhorst v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>In re Eagle Food Ctrs.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>In re Diet Drugs</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gibson v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Zardo v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Beisler v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Watts v. U.S.</td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (scrambled)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>and retirement qualified plan also a 105(c) meeting disability as health</td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Estate of Barnhorst v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>In re Eagle Food Ctrs.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>In re Diet Drugs</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Moore v. Raytheon Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>In re Ullico Inc. Litig.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Albertson’s Inc. v. Comm’r</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>In re Cook</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Findling v. U.S.</td>
<td>N</td>
</tr>
</tbody>
</table>

\(^{122}\) Search done on Nov. 7, 2019. The initial straightforward search on July 9, 2019, produced all irrelevant results, which I did not think to record as a table, and so they are not included. It appears the search algorithm improved between these dates. Search engines are not static, and so rerunning searches can be profitable, though unsettling if you desire consistency.
However, reversing the terms of the query actually improves the results (with some of the key cases still listed first). Finally, scrambling the search query also produces new results (that are irrelevant), while preserving some key cases at the top of the results list. It is hard to know what is going on in Casetext. The results vary based on search term but, regardless of order, the same three relevant cases appeared in the top three positions in each search, and indeed the top five cases (two of which are irrelevant) were the same in all three searches. This suggests a “bag of words” vector approach with some other elements in the search algorithm. Casetext did seem to get nearer to the issues in its results because they were all about retirement plans, but not all the cases were on point for the specific issue of combining a retirement plan with an accident/health plan.

Ravel

Ravel did not produce any relevant results in a straightforward query, but when reversed the results were identical. No results appeared in the scrambled version of the query, which had a problem with starting a query with the term and. Once that was removed, results were similar but not identical to the prior two queries, although without relevant results.

Table 18

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c) as a health and disability plan</td>
<td>Taylor v. Phoenixville Sch. Dist.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Terry v. Bayer Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Regents of the Univ. of Cal. v. Bakke</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Gaworski v. ITT Commercial Fin. Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Schwalm v. Guardian Life Ins. of Am.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Smith v. City of Jackson</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Bogan v. Holland</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Goldstein v. Johnson &amp; Johnson</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Funk v. Cigna Grp. Ins.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Tobin v. Liberty Mut. Ins.</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (backwards)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>plan disability and health a as 105(c) meeting also plan retirement qualified</td>
<td>Taylor v. Phoenixville Sch. Dist.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Terry v. Bayer Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Regents of the Univ. of Cal. v. Bakke</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Gaworski v. ITT Commercial Fin. Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Schwalm v. Guardian Life Ins. of Am.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Smith v. City of Jackson</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Bogan v. Holland</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Goldstein v. Johnson &amp; Johnson</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Funk v. Cigna Grp. Ins.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Tobin v. Liberty Mut. Ins.</td>
<td>N</td>
</tr>
</tbody>
</table>

123. Searches performed on Nov. 19, 2019.
Because of the similarity in results, Ravel appears to be using a “bag of words” vector process.

**Google Scholar**

§70 Google Scholar produced no relevant results with any of the searches or any consistent pattern of results suggesting “bag of words.”\(^{124}\) Citation counts are an important feature of Google Scholar results.\(^ {125}\) So is term frequency.

<table>
<thead>
<tr>
<th>Search Query (scrambled)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>and retirement qualified plan plan also a 105(c) meeting disability as health</td>
<td>Taylor v. Phoenixville Sch. Dist.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Terry v. Bayer Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Bogan v. Holland</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Schwalm v. Guardian Life Ins. of Am.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Goldstein v. Johnson &amp; Johnson</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Gaworski v. ITT Commercial Fin. Corp.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Regents of the Univ. of Cal. v. Bakke</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Smith v. Jackson</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Funk v. Cigna Grp. Ins.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Hunt v. Hawthorne Assocs.</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c) as a health and disability plan</td>
<td>Pilot Life Ins. Co. v. Dedeaux</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Firestone Tire &amp; Rubber Co. v. Bruch</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>McDonnell Douglas Corp. v. Green</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Alexander v. Choate</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Shaw v. Delta Airlines</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Allis-Chalmers Corp. v. Lueck</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Varity Corp. v. Howe</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>L.A. Dep't of Water &amp; Power v. Manhart</td>
<td>N</td>
</tr>
</tbody>
</table>

\(^{124}\) Search done on Nov. 7, 2019.

\(^{125}\) Nevelow Mart, *supra* note 34, at 405, ¶ 28.
### Search Query (backwards)

<table>
<thead>
<tr>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexander v. Choate</td>
<td>N</td>
</tr>
<tr>
<td>Firestone Tire &amp; Rubber Co. v. Bruch</td>
<td>N</td>
</tr>
<tr>
<td>Metro. Life Ins. Co. v. Glenn</td>
<td>N</td>
</tr>
<tr>
<td>Hendrick Hudson Dist. Bd. of Ed. v. Rowley</td>
<td>N</td>
</tr>
<tr>
<td>San Antonio Indep. Sch. Dist. v. Rodriguez</td>
<td>N</td>
</tr>
<tr>
<td>McDonnell Douglas Corp. v. Green</td>
<td>N</td>
</tr>
<tr>
<td>Pilot Life Ins. Co. v. Dedeaux</td>
<td>N</td>
</tr>
<tr>
<td>School Bd. of Nassau Cty. v. Arline</td>
<td>N</td>
</tr>
<tr>
<td>Goldberg v. Kelly</td>
<td>N</td>
</tr>
</tbody>
</table>

### Search Query (scrambled)

<table>
<thead>
<tr>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varity Corp. v. Howe</td>
<td>N</td>
</tr>
<tr>
<td>Pilot Life Ins. Co. v. Dedeaux</td>
<td>N</td>
</tr>
<tr>
<td>Metro Life Ins. Co. v. Mass.</td>
<td>N</td>
</tr>
<tr>
<td>Firestone Tire &amp; Rubber Co. v. Bruch</td>
<td>N</td>
</tr>
<tr>
<td>Alexander v. Choate</td>
<td>N</td>
</tr>
<tr>
<td>Metro. Life Ins. Co. v. Glenn</td>
<td>N</td>
</tr>
<tr>
<td>McDonnell Douglas Corp. v. Green</td>
<td>N</td>
</tr>
<tr>
<td>Goldberg v. Kelly</td>
<td>N</td>
</tr>
<tr>
<td>St. Mary’s Honor Ctr. v. Hicks</td>
<td>N</td>
</tr>
<tr>
<td>Mathews v. Eldridge</td>
<td>N</td>
</tr>
</tbody>
</table>

¶71 Google appears not to have a “bag of words” approach. It differs markedly from other legal search tools, a fact that should be stressed with students.

**Caution and Summary**

¶72 The reader should be cautioned about generally discounting the merits of some of these search engines based on the foregoing. Other techniques or different queries may produce more relevant results than in this illustration, which is not a proper study. Obviously, many more test searches would need to be run to determine definitively what is going on with each major vendor, but the examples above reveal stark differences. It appears that Westlaw Edge is the most prone to use the techniques based on “bag of word” vectors we have described herein, followed by Casetext. For Lexis Advance, word order seems to matter, suggesting that it uses other algorithms than have been discussed in this section, and indeed a *terms and connectors* search ultimately produced comparable results to Westlaw. Bloomberg also produced a few similar results among the three searches. Ironically, the best search for Bloomberg was when we scrambled the terms in the query. Fastcase produced similar results between searches, perhaps suggesting “bag of words,” but because it could not parse 105(c), none of those similar results were relevant. Ravel produced no relevant results, but identical results for forwards and backwards queries, and similar results for a scrambled study. For Google Scholar, no conclusions
can be drawn because of irrelevant results, at least for the straightforward test query. Reversing and scrambling the results also did not yield similar results that would have suggested a “bag of words” vector approach. It is important to stress that Google Scholar appears to operate differently from other legal search engines. Google and Lexis Advance differ most notably from the other search engines that seem to rely on a “bag of words” approach.

¶73 Besides the issue of natural language processing, there is the issue of my frustration at rerunning searches on the same service on different dates (even sometimes a few hours later) only to get different results. I have documented one instance of that with Lexis Advance in appendixes A and B (especially when compared with Table 12). Similar discrepancies are noted with Casetext.126 There were also things I didn’t document, such as Lexis Ravel graphs radically changing between iterations of running the same search.127 This is another affirmation about the instability of our search systems.128 How can the legal profession bestow cognitive authority when not only are there such wide variations between systems, but the results of any search on any given day can change? It may be that we simply don’t have another option. Finally, I must note that for all their sophistication, existing search technologies do not answer the question of whether retirement and disability plans can be successfully combined. The answer to that actually requires human reading and analysis of the relevant cases—so much for AI solving our research problems without us.

¶74 The question of whether iterative searches on a single service or subset of services will replicate the results of searches on all the search services, or at least find a stable basis of precedent, is one that needs to be and can be studied.129

Word Prediction Techniques and Secondary Sources

¶75 Skip-grams and Bayesian algorithms represent only some of the many tools that data scientists use. One variant of the application of such tools is used to predict word searches—which is not only a feature of Lexis, Westlaw, and Bloomberg, but is familiar on Google searches and typing on text messaging.130 However, Lexis and Westlaw have taken word prediction to a new level, by predicting not only search terms or phrases but also applicable questions, documents, and general sources. In figure 3 is a response from searching in Lexis’s search bar to the single term copyright.

¶76 What is interesting is how Lexis ties in secondary sources as well as search terms and Suggested Questions. Lexis’s strength is in rich secondary materials. However, the results are somewhat troubling about why the particular volume and sections of Nimmer on Copyright are selected with so little input from the user, other

126. See supra note 122.
127. See infra figure 9 for example of Lexis Ravel graphic.
128. The Nevelow Mart study confirms this by showing that approximately 40% of results in each service are unique. See supra notes 34 through 37 and accompanying text.
129. See supra note 43.
130. A typical tool used would be Interactive Query Refinement, a variant of Automatic Query Expansion. See Claudio Carpineto & Giovanni Romano, A Survey of Automatic Query Expansion in Information Retrieval, 44 ACM COMPUTING SURVEYS, Jan. 2012, at 1, 8. Google Suggest is an example of such a feature. Id.
**Figure 3**
Lexis Search Bar with Predictive Response to Search Term
(Reprinted from LexisNexis with permission. Copyright 2020 LexisNexis. All rights reserved.)

<table>
<thead>
<tr>
<th>Sources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nimmer on Copyright</td>
<td>i</td>
</tr>
<tr>
<td>Nimmer on Copyright - Index</td>
<td>i</td>
</tr>
<tr>
<td>ALI CLE Course of Study - Trademarks, Copyrights &amp; Unfair Competition</td>
<td>i</td>
</tr>
<tr>
<td>A Practical Guide to Copyright Law in the Digital Age (MCLE)</td>
<td>i</td>
</tr>
<tr>
<td>Associate's Guide to the Practice of Copyright Law</td>
<td>i</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Documents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>copyright act</td>
<td></td>
</tr>
<tr>
<td>4 Nimmer on Copyright § 13.05</td>
<td></td>
</tr>
<tr>
<td>4 Nimmer on Copyright § 13.03</td>
<td></td>
</tr>
<tr>
<td>4 Nimmer on Copyright § 13.01</td>
<td></td>
</tr>
<tr>
<td>2 Nimmer on Copyright § 7.16</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Legal Phrases</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>copyright</td>
<td></td>
</tr>
<tr>
<td>copyright infringement</td>
<td></td>
</tr>
<tr>
<td>copyright law</td>
<td></td>
</tr>
<tr>
<td>copyright protection</td>
<td></td>
</tr>
<tr>
<td>copyright owner</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Suggested Questions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the elements of copyright infringement?</td>
<td></td>
</tr>
<tr>
<td>What is the burden of proof for copyright infringement?</td>
<td></td>
</tr>
<tr>
<td>What is the statute of limitations for copyright infringement?</td>
<td></td>
</tr>
<tr>
<td>What is the definition of copyright owner?</td>
<td></td>
</tr>
<tr>
<td>What is the definition of common-law copyright?</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4**
Lexis Search Bar with Predictive Response to “copyright fair use transformative”
(Reprinted from LexisNexis with permission. Copyright 2020 LexisNexis. All rights reserved.)

<table>
<thead>
<tr>
<th>Legal Phrases</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>fair use and transformative</td>
<td></td>
</tr>
<tr>
<td>copyright fair use</td>
<td></td>
</tr>
</tbody>
</table>
than the single term, *copyright*. There is simply going to be a lot of noise or irrelevant material in our results.

§77 When I add more terms, “copyright fair use transformative,” the results are narrower, but they now are limited to *Legal Phrases* (see figure 4).

§78 If we select “fair use and transformative,” it leads us to a lot of law review articles. By filtering, for *Secondary Material*, we can get back to *Nimmer on Copyright*, in this instance to volume 4 and § 13.05, which has some quite persuasive material on the matter. Therein lies the problem. The process fails to emphasize that some material is more “cognitively” authoritative (or persuasive as recognized by the profession) and relevant than others. The user has to discern through filtering successfully where that material might lie.

§79 Lexis should be applauded. It has taken steps, using predictive techniques, to try to keep secondary material as part of the search inquiry, even in the age of the single-search box and natural language processing, but the tool needs refinement with more complex search inquiries. Nonetheless, perhaps there is hope for secondary material after all. Of course, this is in Lexis’s interest. Secondary materials is where it really shines. The only problem is that indexes (and human intermediation) get circumvented in the process. Appealing to an index will always be last, if thought of at all.
Advancement of these technologies is critical, especially if indexes are going unused as a method of access and are being replaced by the single-search box. As I mention in the introduction, many, if not most, questions are subject in nature and need different tools, or at least methods, than in a known-item search (which favor online search queries).

Not to be outdone, Westlaw Edge also uses predictive searching to suggest a variety of material and searches (see figure 5).

Furthermore, adding words to Westlaw Edge’s search bar actually yields better results for searches, in this instance “copyright and transformative use” (see figure 6).

The algorithms for predictive searching may work on the same principles as the prediction of context words with word2vec’s Skip-gram with Negative Sampling. Probabilities are calculated using Bayesian techniques (which provides for

---

131. See supra ¶¶ 41–44.
training by humans), and the leading contenders are set forth.\textsuperscript{132} The use of secondary sources (in this case, law review articles) is a step toward preserving the role of secondary materials. What would really be great is the appearance of \textit{American Law Reports} or a treatise.

\%84 Rather than suggesting particular resources, Bloomberg will suggest \textit{Practice Centers} based on predictive search terms (see figure 7).

\%85 More specific terms, like “copyright transformative use,” do not produce suggestions. What is important to note is that the whole process bypasses indexes and tables of contents—human-intermediated access tools. It is a shift in how cognitive authority is accessed.

\textbf{Teaching Search Inquiry in the Age of Natural Language Processing}

\%86 Teaching natural language processing may be overlooked, but it is necessary to make students aware of the diversity of results between services, some of the shortcomings of the services, and things they can do to improve results.

\textbf{Differences in Results}

\%87 Because of the manifold variables in natural language processing, it is imperative that students be aware of the differences that various services are likely to provide. Fortunately, Nevelow Mart has provided a ground-breaking study that ought to be read by students to facilitate their understanding of the different and unique results that the various services provide.\textsuperscript{133} Given the current popularity of word2vec (and chances of its use in legal databases), the importance of the range of context words (or the $k$ variable in capturing results ought to be stressed as affecting outcomes. Shorter ranges yield terms that are similar to the target term, and larger ranges are more inclined to pick up the broader topic.\textsuperscript{134} Unfortunately, we do not know what search engines are utilizing. It is possible that longer search inquiries may take better advantage of search algorithms with larger context ranges (as in my examples of the dual-purpose health-disability qualified retirement plans), but this is still speculation (and the opposite occurs with word prediction in the previous section). In my example of case law research, Westlaw Edge performed particularly well in this situation. Furthermore, it seemed to be able to parse a more general term, “qualified retirement plans,” for the many other particular retirement plans that would fall within this umbrella. This is probably because of the use of word vectors or embeddings, which were in evidence by the reversing and then scrambling of the search terms. It even identified the problem as a tax problem without the word \textit{tax} or complete citations to tax code sections appearing in the search query. For Lexis Advance, we had to revert to terms and connectors (and the use of terms within the umbrella of “qualified retirement plans”) to get adequate performance. Rerunning the search months later also improved results, and indeed running searches again even after a few days provided new results.\textsuperscript{135} This does not produce confidence in stability of results, but perhaps this is a method of searching.

\begin{itemize}
\item \textsuperscript{132} See supra \%48–50.
\item \textsuperscript{133} See Nevelow Mart, \textit{supra} note 34.
\item \textsuperscript{134} See JURAFSKY & MARTIN, \textit{supra} note 110, and accompanying text.
\item \textsuperscript{135} Compare Lexis “scrambled” searches in appendixes A and B. Compare also original Lexis search in Table 12 with appendixes A and B.
\end{itemize}
that should be taught. “If you don’t like what you get, rerun the exact same search again.”

¶88 Certain natural language processing activities, such as classification, are useful in determining negative and positive citations to a work, and may use Naive Bayes classification. What is important to understand is that there are human trainers that help teach the algorithms correct classification. Thus, the human element is not totally divorced from natural language processing. The reputation of services like Thomson Reuters Westlaw and LexisNexis for employing human editors is thus a real benefit. Do services like Bloomberg, Fastcase, Casetext, Ravel, and even Google have legal experts helping train their systems? There is an alternative to in-house legal experts, and that is to track users’ (especially attorneys’) responses to make search results more precise. Without humans, the different search tools may be limited in what they can accomplish. The important thing is to recognize the differences.

Lengthening Search Inquiries

¶89 It is uncertain how long and detailed a query should be. In some circumstances, for natural language processing, having a rich and lengthy search query can greatly improve performance (in terms of precision or recall).136 When a user query contains multiple topic-specific keywords that accurately describe his information need, the system is likely to return good matches; however, given that user queries are usually short and that the natural language is inherently ambiguous, this simple retrieval model is in general prone to errors and omissions.137 For Google the average search query is 2.3 words.138 Maybe this is why search tools try to lengthen searches with predictive terms and automatic query expansion. Students should be challenged to lengthen their searches. It may also be helpful to repeat key terms to give them more weight.139 As an alternative to repeated terms, Westlaw has introduced the “^” as a way to emphasize terms.140

Recognizing Biases of Search Algorithms

¶90 Nevelow Mart lists the following necessary attributes of search algorithms that may bias results and contribute to algorithmic opacity:

- prioritization ("emphasiz[ing] . . . certain things at the expense of others"; like relevance ranking);
- classification (putting an "entity [in a] constituent . . . class"; data training may import human biases);
- association ("marks relationships between entities"); and
- filtering, which "includes or excludes information according to various rules or criteria."141

---

136. See supra ¶¶ 53–74. Tables 11–19 illustrate a fairly detailed search phrase.
137. Carpineto & Romano, supra note 130, at 1.
138. Id.
139. See supra note 69.
141. Nevelow Mart, supra note 34, at 394–95, ¶ 12 (citing Nicholas Diakopoulos, Algorithmic Accountability: Journalistic Investigation of Computational Power Structures, 3 DIG. JOURNALISM 398, 399 (2015)).
¶91 Nevelow Mart calls for disclosing how these factors are used for “algorithmic accountability.”142 Helping students see these categories of potential bias is something worthy of classroom discussion.

Recognizing Out-of-Context Matches

¶92 There are a variety of problems with natural language search inquiries that students would do well to recognize:

- Polysemy (or the multiple meanings of words, such as “organization”);
- Relationships based on order (“man bites dog,” rather than “dog bites man”);
- Out-of-phrase terms (failure to treat a phrase as a single unit, such as “North Atlantic Treaty Organization”);
- Secondary topic key words (“Roth IRA” versus “IRA”);
- Noncategorical terms (“fair use” is an instance of an exception under both copyright law and trademark law).143

¶93 To this list, I would add my own frustration in doing online search queries of the codes of all 50 states on Lexis Advance and Westlaw Edge to determine the states that had repealed or otherwise done away with the “rule of perpetuities.” All my results affirmed the rule in state after state. I conclude that besides being difficult to do natural language searching in codes, it is difficult to prove a negative (or a repealed statute). Only after locating a secondary source on the topic was I able to start to find states that had done away with the rule.

Secondary Sources and Codes

¶94 The problem is that many searches should start with secondary sources (or codes),144 but the interfaces of the major vendors place them in a secondary position. Lexis Advance and Westlaw Edge (and to some extent Bloomberg Law) are to be applauded by suggesting the user types in the search box sources that could be used. But these services are still primitive: the more you type in the search box, the less you get. Compared to a good index in a secondary source, the insufficiency is noteworthy.

¶95 Most of the time, to get to secondary sources, users have to search and then select filters that bring secondary sources to the forefront. This totally bypasses any use of human-intermediated indexes. Students need to be instructed both to get into the habit of “drilling down” to secondary sources and, in many instances, to search out the index for initial queries. And then there is the question of what to do about statutory codes. My own experience in natural language searching on codes has limited utility. Indexes and tables of contents are still key and need to be taught at every opportunity with students.

Reiterative Searching

¶96 Often, it is not the initial search query that produces the best results, but what the researcher does with those results that ultimately finds the best cases or authority. This is illustrated in figure 8.

142. See Nevelow Mart, supra note 34, at 395, ¶ 14.
143. Carpineto & Romano, supra note 130, at 5.
144. See Callister, supra note 28, at 24–34.
Teaching students to find similar authority is thus vital when initial search queries produce such varying results. There are a variety of techniques for finding similar authority. Westlaw Edge or Lexis Advance headnotes can lead to similar points of law, if the particular point in question has been assigned to a unique headnote. As with search queries, the differences between headnotes is significant.
between the two vendors, and so using them is more art than science. Users may also search headnotes and particular terms at the same time, yielding the greatest opportunity for stumbling on similar cases. Citation analysis can also lead to more relevant cases. Both the cases cited in a case (from the table of authorities), and the cases that cite a relevant case can lead to stronger authority. Finally, understanding key cases is important. Consequently, tracking commentary (like an American Law Reports annotation or an encyclopedia, such as California Jurisprudence) that has cited a case may uncover better authority—the ones practitioners rely on.

¶98 One technique for reiterative searching is to follow up a natural language search with a terms and connectors search, just to compare. A particularly useful tool is having identified a key term or phrase to search within results using an “atleastN” command. But the use of commands takes us out of natural language processing and this article’s quest to explore how responsive such systems are to legal parlance.

Search Result Visualizations

¶99 Data visualizations, like those provided by Lexis and Fastcase, often can be used to winnow key authority to seminal cases on an issue. Figure 9 shows my search of federal case law for cases with “fair use” and “transformative.” While the first case that appears is Fox News Network v. TV Eyes, I can readily discern by rolling over the large circle in the Supreme Court segment of the Ravel View chart that Campbell v. Acuff-Rose Music is seminal to citations about fair use and transformative use. Thus, while my natural language search may not have produced the most relevant or seminal cases first, the Lexis Advance Ravel View visualization function quickly got me the information I needed.

¶100 Likewise, with Fastcase, a terms and connectors search for “copyright and ‘fair use’ and transformative” yields Fox News Network v. TV Eyes as the first hit, but when I switch to interactive timeline, with the vertical axis set to court level, I can readily see that Campbell v. Acuff-Rose Music is the key Supreme Court precedent, heavily cited, on the issue (see figure 10). Interestingly, Westlaw Edge (lacking the visualization tools) in a natural language search of “fair use transformative” produced Campbell v. Acuff-Rose Music as its third entry (neither Fastcase nor Lexis produced the case in top results). The point is the means to get to the seminal case, preferably a Supreme Court case, varies from service to service, and this needs to be taught to students.

¶101 Thus, visualizations can be a powerful tool in discerning the most relevant authority on an issue after conducting a primary search.

Conclusion

¶102 By undertaking this article, I assumed a large risk—chiefly because I don’t know how the major legal search engines employ natural language processing and

---

145. See Hook & Mattson, supra note 111 (study on difference of headnotes between Westlaw and Lexis).
146. Rerunning a search at a later date or time may produce different results and a different graphic.
147. A natural language search of the same terms yielded too many non-copyright cases.
AI.\textsuperscript{148} I have made some educated guesses based on the literature about natural language processing in hopes of gleaning some understanding of what may be going on with searches across the various vendors. I have even tried to find evidence for tools, such as “bag of words,” by reversing and scrambling search inquiries to see what results were produced.\textsuperscript{149} Besides search tools (and comparison of documents for similarity), I have introduced Bayesian classifiers that may be used for a variety of tasks, such as detecting when cases are distinguished or not followed (or classification of headnotes by topic or Key Numbers).\textsuperscript{150} Are any of these tools used? I have no way of knowing, other than statements by the major services that they use AI.\textsuperscript{151}

¶103 Vendor promotional material proclaiming, \textit{You Ask a Question . . . Lexis Answers Understands It},\textsuperscript{152} is marketing research services that are intelligent. However, as we understand more about natural language processing, some of the “magic” goes away, and we see algorithms at work. If the major vendors are using neural networks\textsuperscript{153} to find through regression optimum weights for vectors or embeddings, this may be one step closer to AI (and is a part of \textit{machine learning}).\textsuperscript{154} But this really is the application of the brute force of computing power. The problem of such applications of “brute force” is that they command vast computing resources and expense, perhaps making themselves available to only the most prosperous sector of the legal profession.\textsuperscript{155}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{FastcaseTimeline.png}
\caption{Fastcase Timeline for Search of “Copyright and ‘Fair Use’ and Transformative”}
\end{figure}

\textsuperscript{148} See supra note 8.
\textsuperscript{149} See supra ¶¶ 53–74.
\textsuperscript{150} See supra ¶¶ 48–50.
\textsuperscript{151} See supra notes 8, 30, 47.
\textsuperscript{152} See LexisNexis, \textit{supra} note 30.
\textsuperscript{153} See supra ¶ 47.
\textsuperscript{154} Jason Brownlee, \textit{Logistic Regression for Machine Learning}, \textit{Machine Learning Mastery} (Apr. 1, 2016), https://machinelearningmastery.com/logistic-regression-for-machine-learning/ (last visited Aug. 5, 2020) (“Logistic regression is another technique borrowed by machine learning from the field of statistics.”); \textit{Wikipedia}, \textit{supra} note 85 (“Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead.”).
\textsuperscript{155} See Jones, Kalantry & Glover, \textit{supra} note 4, at 18 (“Unless the cost of compute
¶104 The larger concerns recited in the beginning of this article were stability of the law, opacity, and de-emphasizing or even replacing secondary authority (the great treatises), thereby affecting cognitive authority. Stability of law is a concern because of inconsistent results and the corresponding fact that very little of what attorneys are citing in briefs appears to make it into court decisions. Inconsistent results among services is compounded by the fact that most attorneys may not have access to all the services, nor can clients pay for such prodigious use of research time. However, assuming access to the right research services, there are techniques to take initial results and find seminal cases through headnotes or visualization. But the first round of searching can produce widely different results. Following up with additional steps to the first round of results can improve performance.

¶105 Is opacity among search engines a threat? We have turned over the stability and structure of law to data scientists. However, as the author of Weapons of Math Destruction points out, recent history is rife with examples of unintended and harmful consequences from algorithms. The lack of stable responses to search queries is a threat, but there may be others, such as biases yet unseen. Nevelow Mart has suggested that the case classification systems used by West and Lexis are possibly biased for being based on the 19th century Langdellian conceptions of the law. Nevelow Mart even speculates about the possible bias (or at least the differences in results) “from the very different list of secondary sources in Westlaw and Lexis Advance [that] are baked into their respective search results . . . .” Secondary sources may well be part of the algorithmic equation when searching case law in the two dominant services.

¶106 Finally, there is the question about cognitive authority and how it might change. To answer the question, I have introduced Deibert’s holistic ecological model of media theory that eschews technical determinism. Within its rings, language reigns supreme as a technology through time. That humans are driven to have machines process and even understand language, including in such technical fields as law, is inherent in our nature. The introduction of natural language processing is already affecting law’s cognitive authority, but its historical parameters (like 19th century classification systems) will also affect natural language processing and its use. At least, that is the prediction of Deibert’s model. “Langdellian” conceptions may survive the migration to new cognitive authority.

¶107 In the introduction, I was also concerned about the demotion of secondary sources because a single-search box tends to default to case law, and finding secondary material requires filtering. This is somewhat alleviated by the tools of Lexis Advance and Westlaw Edge that suggest secondary materials as part of the predictive searching, but these services are still primitive, and when more text is entered drastically decreases, experiments will grow too large to be affordable by anyone but the US or Chinese governments. Compute is the correlation between the amount of computing power used to train AI and the power necessary for the resulting AI model. Id. supra ¶¶ 75–85.

156. See supra ¶¶ 75–85.
157. See Bennardo & Chew, supra note 35.
158. See supra ¶¶ 96–101.
159. See O’Neil, supra note 2.
160. See supra note 34.
161. Nevelow Mart, supra note 34, at 419, ¶ 55.
162. See supra notes 14 through 16 and accompanying text.
163. See supra ¶¶ 75–85.
into the query, secondary material seems to disappear, leaving the user to rely on post-search filtering. Finally, the indexes and tables of contents are lost in the single-search box world. That connection to human-intermediated information is disappearing.¹⁶⁴

¶108 There are a number of reasons why we cannot yet rely upon legal search engines like we do with Google or Amazon Alexa, at least with respect to natural language processing. The legal search services are turning to the single-search bar, but their natural language algorithms appear to be based on a “bag of words,” with scarce evidence of syntax playing any significant role.¹⁶⁵ For the most part, by treating words as vectors, we are relying on proximity—whether terms in the same document or within a “context window”—to accord with meaning. I believe legal discourse to be too subtle to be boiled down to proximity of words, regardless of the method of doing so. Perhaps more important, the role of secondary authority (the great treatises) is often subjugated in search processes, especially with respect to access through indexes and tables of contents.¹⁶⁶ Indeed, the use of the latter is dying. Many up and coming legal research services do not even have secondary materials. Add to that the difficulty searching legal codes. Often, commentary and codes should be the starting point of legal research rather than an afterthought.

¶109 In conclusion, the best we librarians can do in the face of uncertainty is to teach our users about the limitations of these systems, disillusioning them of computer intelligence doing the work for them—at least for now. If anything, AI is a tool and, one day perhaps—assuming a humanistic techno-central vision—a partner. Perhaps the day will soon come when law firms will list IBM’s Watson as a partner, and we will be able to ask questions of legal search engines like we do with Google and Alexa. But that day is still a ways off. Whatever may happen, the profession’s shared cognitive authority is shifting to the algorithm—it is in it that we will entrust our future.

¹⁶⁴. See id. Hopefully West’s reliance on the human-intermediated Key Number System will counter this trend. See supra notes 47–48 and accompanying text.

¹⁶⁵. There is evidence that sometimes word order does matter in some of the searches conducted, see supra ¶¶ 53–74, because not every search engine produced identical results regardless of search order.

¹⁶⁶. See supra ¶¶ 75–85.
### Appendix A

**Lexis Advance**

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c) as a health and disability plan(^{167})</td>
<td>Gibson v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Wright v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Paul v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Fr. v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Chernik v. Comm’r</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>ABA Ret. Funds v. U.S.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Enloe v. Comm’r</td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (backwards)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>plan disability and health a as 105(c) meeting also plan retirement qualified</td>
<td>No cases</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (scrambled)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>and retirement qualified plan also a 105(c) meeting disability as health</td>
<td>In re Disney ERISA Litig.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>In re Alpha Natural Res.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>U.S. v. Martinez</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>ABA Ret. Funds v. U.S.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Antioch Co. v. Morgan</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Zardo v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gentile v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Kifafi v. Hilton Hotels Ret. Plan</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Thomas v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Thomas v. Comm’r(^{168})</td>
<td>Y</td>
</tr>
</tbody>
</table>

---

168. The results are both to the United States Tax Court Opinion, T.C. Summary Opinion 2007-110 (Jun. 28, 2007).
## Appendix B

### Lexis Advance

<table>
<thead>
<tr>
<th>Search Query (federal cases)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualified retirement plan also meeting 105(c) as a health and disability plan(^{169})</td>
<td>Gibson v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Wright v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Paul v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Fr. v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Chernik v. Comm’r</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>ABA Ret. Funds v. U.S.</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Enloe v. Comm’r</td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (backwards)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>plan disability and health as 105(c) meeting also plan retirement qualified</td>
<td>No cases</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Query (scrambled)</th>
<th>Case</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>and retirement qualified plan plan also a 105(c) meeting disability as health</td>
<td>Gordon v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Caplin v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r (6th Cir.)</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Gibson v. U.S.</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Berman v. Comm’r (Tax Ct.)</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Hall v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>West v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Kelter v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Dorroh v. Comm’r</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Armstrong v. Comm’r</td>
<td>Y</td>
</tr>
</tbody>
</table>

---

\(^{169}\) Search run Nov. 18, 2019. Even identical searches run the same day may have different results on Lexis Advance.
The Shadow Code: Statutory Notes in the *United States Code*∗

Shawn G. Nevers** and Julie Graves Krishnaswami***

*This article examines the history, creation, and purpose of statutory notes in the United States Code. It also explores the challenges statutory notes present in the research process and offers guidelines and best practices for researching and teaching about this often overlooked aspect of the United States Code.*

Introduction ......................................................... 214
Editorial Versus Statutory Notes ........................................ 215
History of the *United States Code* ..................................... 217
Creation of Statutory Notes ............................................ 222
  Positive Law Titles—OLRC’s Forced Hand ............................. 223
  Nonpositive Law Titles .............................................. 227
  Extent of the Problem ............................................... 230
The Shadow Code .................................................... 236
  Inconsistency ...................................................... 238
  Placement ......................................................... 239
  Legal Research Providers ............................................ 240
Research ............................................................... 244
  General Principles .................................................. 245
  Secondary Sources .................................................. 245
  Statute Compilations ................................................ 246
Searching ......................................................... 249
Index ............................................................. 250
Improving Research Access .......................................... 251
Best Practices for Teaching Law Students About Statutory Notes .......... 252
Conclusion .......................................................... 255

** Deputy Director, Howard W. Hunter Law Library, BYU Law School.
*** Head of Instruction and Lecturer in Legal Research, Lillian Goldman Law Library, Yale Law School. Special thanks to Law Revision Counsel, Ralph Seep, and Deputy Law Revision Counsel, Robert Sukol, for their important insights and feedback. Thanks to David Armond, Mari Cheney, Iantha Haight, Annalee Hickman, Susan Nevelow Mart, Ingrid Mattson, Fred Shapiro, and Leslie Street, as well as participants of the Boulder Conferences on Legal Information at the University of Texas School of Law (2017) and the University of Maryland, Francis King Carey School of Law (2018), and participants of the Utah Legal Research Instruction Best Practices Workshop for their helpful feedback on earlier versions of this paper. We also thank Jacob Munyan and Sophie Laing for their research assistance.
“A provision of a Federal statute is the law whether the provision appears in the Code as section text or as a statutory note.”

Introduction

\[1\] In April 2017, the D.C. Circuit decided what some commentators called the “Snakes on a Plane” case.\[2\] The issue presented in that case was whether the Lacey Act’s shipment clause prohibited the transportation of reticulated pythons and green anacondas from one state to another.\[3\] In dismissing one of the government’s arguments, the court cited Congress’s definition of “continental United States.”\[4\] At first glance, there’s not much to see here. Courts cite definitions from the United States Code all the time. But here the D.C. Circuit did not cite a United States Code section to define “continental United States.” It cited a statutory note.

\[2\] Statutory notes are provisions of law placed after the text of a United States Code section. They exist throughout the United States Code and are valid law despite their location in the Code. The first eight sections of the United States Code set forth definitions of words such as “county,”\[5\] “vessel,”\[6\] and “vehicle”\[7\] that apply to all acts of Congress. Yet the definition of “continental United States” is relegated to a statutory note.\[8\] Despite its location in the notes, the definition of “continental United States” is law: it was passed by Congress and signed by the President, and should be treated and relied upon like any other provision of law.\[9\] The problem is that it doesn’t show up where most laws do, causing bewilderment and confusion even for experienced researchers.\[10\] And “continental United States” is not alone. Many statutes end up in notes to the United States Code rather than in Code sections. The inclusion of laws as statutory notes has several consequences. First,


4. Id. at 1140.


8. 1 U.S.C. § 1 note. One author has called statutes that affect the meaning of later-passed statutes, including 1 U.S.C. § 1, submarine statutes. Christian Turner, Submarine Statutes, 55 Harv. J. Legis. 185, 190 (2018). A note to a submarine statute is an even more troubling problem, raising the possibility of stealth submarine statutes.


10. See Schwier v. Cox, 340 F.3d 1284, 1288 (11th Cir. 2003) (explaining that the district court mistakenly noted that “although section 7 was part of the Privacy Act that ‘was passed into law as Public Law 93-579,’ the fact that section 7 ‘was never codified, and appears only in the ‘Historical and Statutory Notes’ section of the United States Code,’ made section 7 a mere ‘historical footnote to the Privacy Act of 1974 [which] Congress has never reflected any intention of [codifying]’”); Michael J. Lynch, The U.S. Code, the Statutes at Large, and Some Peculiarities of Codification, 16 Legal Reference Servs. Q. 69, 80 (1997) (”It is astonishing that laws of general significance . . . should be found in the United States Code only in the notes.”).
because these notes are separate from the main text, the *United States Code* is not a “unified and comprehensive text.” Second, this separation of the notes and main text makes researching federal statutory law more difficult. Because this system is not intuitive and contemporary researchers often assume that all current federal statutory law is found in the *United States Code*, researchers are likely to struggle finding law in statutory notes. Novice researchers may find it nearly impossible to find law that exists in statutory notes.

§3 Other commentators have discussed statutory notes too, but this article goes beyond the existing literature in a few ways. First, this article provides a more detailed explanation of why statutory notes exist. Second, it provides a fuller picture of the extent to which statutory notes exist and play a role in the practice of law. Third, this article examines other factors that obscure statutory notes, including legal research providers’ presentation of these notes and how this affects a legal researcher’s ability to discover statutory notes effectively in the research process. Finally, this article presents best practices for researching and teaching statutory notes.

**Editorial Versus Statutory Notes**

§4 At the outset it is important to establish that there are different types of notes in the *United States Code*. The official version of the *United States Code*, prepared by the Office of the Law Revision Counsel (OLRC), contains a section of notes following the statutory text of each section of the Code. There are two main types of notes that are included—editorial notes and statutory notes. Editorial notes are written by the OLRC and assist researchers in understanding a code section. Editorial notes “provide information about the section’s source, derivation, history, references, translations, effectiveness and applicability, codification, defined terms, prospective amendments, and related matters.” Revision notes, for example, are a type of editorial note that briefly explain what changes were made to a section of the Code by certain amendments. This provides researchers with a useful way to narrow in quickly on which amendments to focus on in legislative history research. Another category of editorial notes, codification notes, provide information about a section’s relationship with other sections. While useful research tools, editorial notes are not law.

12. See Tobias A. Dorsey, *Some Reflections on Not Reading Statutes*, 10 Green Bag 2d 283 (2007) (“So we read the Code first. If it is clear, we read nothing else. If it is not clear, we look at the old Code. We no longer ask what Congress wrote; we ask what the Code says.”). Even researchers who understand the evidentiary status of the *United States Code* may still forget to check the underlying law. See William Baude & Stephen E. Sachs, *Grounding Originalism*, 113 Nw. U. L. Rev. 1455, 1480 (2019).
16. Id. at https://uscode.house.gov/detailed_guide.xhtml#editorial [https://perma.cc/24MS-3CJD].
17. See id.
Statutory notes, on the other hand, are law. The OLRC defines them as “provisions of law that are set out as notes under a Code section rather than as a Code section.”\(^\text{18}\) Unlike editorial notes, statutory notes have been passed into law and have the force of law despite not showing up as independent code sections.\(^\text{19}\) The decision to place these provisions of law in statutory notes is an editorial decision made by the OLRC—a decision sometimes forced upon it—but their existence and location as statutory notes rather than code sections does not affect their meaning or validity.\(^\text{20}\) Generally, statutory notes follow editorial notes, but the break is not clearly distinguished in the Code and can be difficult to spot. Statutory notes do not have a consistent heading within the Code and are generally identified by a heading that names the public law of origin.

Figure 1 shows 28 U.S.C. § 1350, commonly known as the Alien Tort Statute, and the beginning of its notes section. Following the text of § 1350 and its credits are editorial notes labeled “Historical and Revision Notes.” Next appears a statutory note labeled “Torture Victim Protection.” While the heading does not specify this as a statutory note, it does provide a unique heading related to the title of the public law—here the Torture Victim Protection Act of 1991. An even better indicator that

---

18. Id. at https://uscode.house.gov/detailed_guide.xhtml#statutory [https://perma.cc/VF7C-W46V].


this is a statutory note is the phrase “Pub. L. . . . provided that:” followed by a quotation of statutory language.\textsuperscript{21}

\textsection{7} If the distinction between editorial and statutory notes isn’t confusing enough, commercial versions of the United States Code, which most researchers use, add another wrinkle. These versions of the Code include an additional category of notes in the form of annotations consisting of citations to case law, legislative history documents, administrative regulations, and secondary sources. These notes appear near the editorial and statutory notes from the OLRC, and it is difficult to see where one ends and another begins unless a researcher is extremely familiar with the content of each. In fact, in some instances commercial publishers modify the OLRC’s notes, making it even more difficult for researchers to understand who—the commercial code editors or the official codifiers at the OLRC—is providing the information.\textsuperscript{22}

\textsection{8} It’s fairly obvious that editorial notes and commercial publishers’ notes benefit researchers. Although these editorial notes are not law, they point to law and other helpful sources. Statutory notes, on the other hand, are law—so why aren’t they contained in sections of the United States Code? The answer is multifaceted and more complicated than necessary. A partial answer is the existence of positive law and nonpositive law titles in the United States Code. “Positive law,” in the context of the United States Code, means that a title of the Code has itself been enacted as a statute and is legal evidence of the law.\textsuperscript{23} Nonpositive law titles, on the other hand, are merely compilations of statutes and are only \textit{prima facie} evidence of the law.\textsuperscript{24} Unfortunately, the current United States Code contains both positive and nonpositive law titles. To better understand the distinction between positive and nonpositive law, and to understand why this distinction exists and how it affects statutory notes, it’s useful to examine briefly the history of the United States Code.

\begin{center}
\textbf{History of the United States Code}
\end{center}

\textsection{9} For the first many years of the United States’ existence, there was nothing like the official United States Code we have today. Bills were passed into law and were then published as slip laws or session laws in newspapers or occasionally as their own collections.\textsuperscript{25} The \textit{Statutes at Large}, which began a more systematic publication

\textsuperscript{21} There are some slight variations on this, including “Pub. L. . . . provided:” or “Act . . . provided that.”

\textsuperscript{22} Lexis, for example, provides editorial and statutory notes in a section labeled “Annotations,” along with case notes. Westlaw adds citations to legislative reports, before other editorial notes begin. \textit{Compare}, e.g, 8 U.S.C. § 1101 (note) with 8 U.S.C.S. § 1101 (note) and 8 U.S.C.A. § 1101 (note).


\textsuperscript{24} \textit{See Positive Law Codification, Off. L. REVISION COUNSEL: UNITED STATES CODE, http://uscode.house.gov/codification/legislation.shtml [https://perma.cc/QVB2-Y7Z2]; Sukol, supra note 23, at 11–12 (“Nonpositive law titles, as such, have not been enacted by Congress, but laws assembled in nonpositive law titles have been enacted by Congress.”)}.

\textsuperscript{25} \textit{See Ralph H. Dwan & Ernest R. Feidler, The Federal Statutes—Their History and Use, 22 MINN. L. REV. 1008, 1009 (1938); Erwin C. Surrency, A HISTORY OF AMERICAN LAW PUBLISHING 100–04
of federal session laws, was first published in 1845. In 1866, President Andrew Johnson “appointed a commission to revise, simplify, arrange, and consolidate all statutes of the United States, general and permanent in nature.” After many years of discussion and work, the Revised Statutes of 1873 were passed into law while simultaneously repealing all laws enacted prior to December 1, 1873, that were “embraced in any section” of the Revised Statutes. Thus, the Revised Statutes of 1873 became legal evidence of the law (that is, actual evidence, stronger than prima facie evidence of the law). Researchers could now turn to the Revised Statutes of 1873 to determine what the law was, rather than resorting to the Statutes at Large or prior acts. While not referred to as such at the time, the Revised Statutes of 1873 are an example of positive law—a compilation that itself was officially passed into law through the legislative process.

The success of the 1873 Revised Statutes, however, was short lived. Almost immediately, Congress began receiving reports of errors in this first Revised Statutes and, in response, appointed someone to prepare a new edition. By 1877, the concept of codifying federal laws had become widespread, and Congress saw the need for a second edition of the Revised Statutes. The second edition was completed in 1874 and published in 1875, with the title Revised Statutes of the United States of America, general and permanent in nature. The new edition included amendments and corrections to the original edition, but it was not error-free. Researchers continued to turn to the Revised Statutes of 1873 for guidance, and Congress was left to work with an incomplete and imperfect code. The process of revising and updating the law continued, and by 1913, the United States Code was published, consolidating all federal laws into a single, comprehensive code.
Congress felt compelled to pass legislation that provided for the appointment of a commissioner to prepare a new edition of the Revised Statutes that would include the amendments passed since 1873. In this act, Congress provided that the new edition of the Revised Statutes would again be “legal and conclusive evidence of the laws.” A year later, however, Congress changed course. Concerned that courts were unable to look back to the original acts because they had been repealed, Congress amended the 1877 act to make the forthcoming edition only prima facie evidence of the law. In other words, the Revised Statutes of 1878 would be a useful compilation of statutes in force since 1873, but it could be rebutted by reference to a discrepancy in the Statutes at Large. It would not be positive law.

At the turn of the 20th Century, there was a move for the creation of an official United States Code that would organize and compile all law that was currently in force by topic. In 1897, Congress authorized the appointment of a commission to revise and codify United States criminal and penal laws. In 1899, this charge was expanded to also include laws surrounding jurisdiction and practice of courts of the United States. And, finally, in 1901, Congress empowered the commission to revise and codify all federal laws of a permanent and general nature. One house report explains that there was “a widespread demand upon the part of the legal profession and of that considerable part of the general community directly interested in the application and enforcement of the law for a clear, perspicuous, and systematic compilation of the laws of the United States.”

After a number of years of work, Congress passed a Criminal Code in 1909 and a Judicial Code in 1911—both as positive law. The Committee on Revision of Laws next turned its sights to revising and codifying all permanent and general law, presenting Congress with the first version of a comprehensive Code in 1919. This draft, along with several later drafts, would have repealed prior Acts and enacted the code as positive law. Each of these drafts was rejected by the Senate after errors were discovered, likely reminding Senators of the error-prone 1873 Revised Statutes to the amendment of every member who may think that any branch of the law needs change.” To remedy this problem, the committees appointed Thomas Jefferson Durant to review the work of the commissioners and eliminate any changes they had made. Durant spent nine months on this work, but errors naturally persisted because the goal of the work changed from a revision that allowed for some change in the law where necessary to a strict restatement of existing law. at 650.

32. 19 Stat. 267 (1877).
33. Id.
34. 20 Stat. 27, ch. 26 (1878); 7 CONG. RECD 1137 (Feb. 18, 1878); 7 CONG. RECD 1376–77 (Feb. 27, 1878).
35. 30 Stat. 58 (1897).
36. 30 Stat. 1116 (1899).
37. 31 Stat. 1181 (1901).
39. 35 Stat. 1088 (1909); 36 Stat. 1087 (1911). Interestingly, Congress passed the Criminal and Judicial Codes as positive law, repealing all prior acts and portions of the Revised Statutes covered by the new codes. A house report accompanying one of the criminal code bills indicates that the Committee considered subject by subject revision and codification as the preferable way to approach the process based on the unsuccessful 1873 Revised Statutes. See H.R. REP. No. 59–3200, supra note 38. Unfortunately, Congress abandoned that approach with the 1926 Code.
41. Id.
utes. With the help of legal publishers, the Committee on Revision of the Laws presented a new proposed Code to Congress in 1926. This bill contemplated enacting the Code into positive law, but with a grace period for errors to be discovered and changed before the Code became positive law. While the House passed this version of the bill, Senators ultimately felt uncomfortable repealing all prior law in favor of the new code and instead opted to make the Code prima facie evidence of the law only. Those in favor of positive law enactment were disappointed but took solace in the fact that positive law codification would happen relatively quickly. How very wrong they were.

The process of positive law codification began in 1940, but the first positive law bills did not pass until 1947, more than 20 years after the creation of the United States Code. Between 1924 and 1947, all titles of the Code were nonpositive law titles. In addition to leaving the Code as merely prima facie evidence of the law, this state of affairs meant that when amending a statute Congress would amend the underlying act, not the Code itself. That began to change when positive law titles

42. William L. Burdick, The Revision of the Federal Statutes, 11 A.B.A. J. 178, 181 (1925) (some 600 errors were discovered, although Burdick argued that only 100 or so were of importance); Dwan & Feidler, supra note 25, at 1021; Whisner, supra note 13, at 552 n.32 (quoting Rep. Roy G. Fitzgerald).
43. Whisner, supra note 13, at 551.
44. S. R/e.sc/p.sc. N/o.sc. 69-832, at 3 (1926) ("It is further desired to call attention to and emphasize the fact that this compilation, unlike many of its predecessors is complete within itself and can be used without the necessity of having to consult the Revised Statutes and the 24 volumes of the Statutes at Large, published since the Revised Statutes."); Lee & Beaman, supra note 40, at 835 (describing the "twilight zone" system that would allow for correction of the Code before it officially became positive law).
45. See 67 C/o.sc/n.sc/g.sc. R/e.sc/c.sc. 12,074 (1926); George Wharton Pepper, Philadelphia Lawyer 164 (1944) (describing Sen. Pepper’s role in devising the plan to pass the Code as evidence of the law and to enact the Code as positive law at a later time). After positive law codification of 13 titles, Charles Zinn, participating in proceedings of the American Association of Law Libraries in 1958, remarks that very few errors had been found in these titles. Charles S. Zinn, Revision of the United States Code, 51 L/a.sc/w.sc L/i.sc/b.sc/r.sc. J. 388, 395 (1958).
46. 67 CONG. REC. 12075 (1926) (Mr. Madden expressed that "the practical effect, of what has been done, if followed up, would result in a very short time in getting a codification of the laws which could be presented as the laws of the land in cases before the court" and judged that it could be done in three Congresses. Mr. Fitzgerald hoped that the Code would become positive law after a few years of experience with it.) Lee and Beaman, writing in 1926, thought that one probable path for the Code would be that if it was shown that a limited number of errors were discovered in the next few years, Congress would pass an Act causing the Code to become law. Lee & Beaman, supra note 40, at 838.
47. See Tress, supra note 11, at 137. Some of this time was considered a testing period in which errors could be identified. See Zinn, supra note 45, at 391. Even in 1941, the Committee on the Revision of the Laws was optimistic that positive law codification for the entire Code would not take long. Preface of United States Code, at ix (1940) ("It is fervently hoped that the program of enacting the Code into positive law, title by title, to improve its present status as merely prima facie evidence of the law, will meet with success in the not too distant future.").
48. There appears to have been some confusion in the House shortly after enactment of the Code, as to whether new amendments should be to the Code or the underlying statute. 69 CONG. REC. 2091 (1928) ("If no one has looked into the question as to just what recognition should be given, I think it ought to be referred to a committee selected to give it study and determine whether or not it would be safe, in bringing in bills to amend statutes, to refer only to the new code of laws and not to the other sources that existed before the new code of laws."). Lee and Beaman, on the other hand, understood that the underlying act, not the Code, would have to be amended. Lee & Beaman, supra note 40, at 837 ("If law is to be changed by orderly process of express repeal and amendment, the basic statutes as appearing in the Revised Statutes and Statutes at Large must be repealed or amended. . . .


were enacted in 1947. Title 17, Copyrights, was one of the titles enacted into positive law in 1947. As part of this process, Congress examined title 17, which was prepared by codifiers, to determine whether it was in accordance with the acts that had been passed by Congress throughout its history. Once Congress felt comfortable that title 17 accurately reflected the law, it enacted the title as positive law and repealed prior provisions of the Revised Statutes and Statutes at Large codified in title 17.49 Now, title 17 would be irrebuttable, legal evidence of the law, and Congress would not need to amend underlying acts but could amend title 17 directly.

¶14 Positive law codification has moved slowly since 1947. To date, 27 of the current 54 titles have been enacted into positive law in the 93 years since the creation of the United States Code.50 And while those 27 titles make up half of the titles in the United States Code, they account for roughly only a quarter of the total pages

[The only result produced by an amendment of the Code would be to make it incorrect as a reproduction of those statutes.]


of the Code. For these titles, what the United States Code says is law. For the other 27 nonpositive law titles, what the United States Code says is only prima facie evidence of the law and can be rebutted by the Statutes at Large. While this distinction is important, it is largely academic. Only rarely does a discrepancy between text in a nonpositive law title and the Statutes at Large occur or significantly influence the outcome of a dispute. But how positive and nonpositive law titles are amended does have a more consistent impact on researchers, as it can affect where statutory provisions are placed in the Code. Understanding the differences in the amendment process for both types of titles can help researchers be better prepared to deal with statutory notes in their research.

Creation of Statutory Notes

¶15 One of the core distinctions between positive and nonpositive law titles in the United States Code is how they are assembled and amended. Positive law titles can be changed only by direct amendment by Congress, and the amendments must reference a change to the title and section of the Code. On the other hand, nonpositive law titles can be arranged or rearranged by the OLRC. These code sections are subsequently amended by reference to the act or public law provisions under revision. The OLRC does not make any substantive changes to the law, but because nonpositive law titles are only compilations of the public laws in force, the OLRC

51. This estimate was arrived at by totaling the number of pages in the 2012 edition of the United States Code for positive law titles and nonpositive law titles. In 2013, former Law Revision Counsel, Peter G. LeFevre, estimated that 70% of the code was not positive law. LeFevre, The United States Code: Its Accuracy, Accessibility, and Currency, ADMIN. & REG. L. NEWS, Winter 2013, at 10, 11. One legislative drafter lamented the “sorry state of the United States Code” in 1961, when 16 of the 50 titles were enacted into positive law. Peacock, supra note 30, at 45. One can only imagine what he would think today with only a modest increase in the number of positive law titles and the number of United States Code volumes ballooning from 10 in 1958 to 34 in 2012.

52. One of which, title 53, is currently reserved and has no content.

53. However, it is likely still shocking to many practitioners who use it frequently. See William Li et al., Law Is Code: A Software Engineering Approach to Analyzing the United States Code, 10 J. BUS. & TECH. L. 297, 300 (2015).


55. While beyond the scope of this article, the positive/nonpositive law distinction also affects other aspects of the research process. Katz and Bommarito developed a model to measure complexity in the United States Code, defining complexity as “the human capital expended by a society when an end user is required to review and assimilate a body of legal rules.” Daniel Martin Katz & M.J. Bommarito II, Measuring the Complexity of the Law: The United States Code, 22 ARTIFICIAL INTELLIGENCE & L. 357, 340 (2014). Interestingly, the five most complex titles of the United States Code, according to Katz and Bommarito’s unnormalized score—which simulates “the complexity of reading and assimilating the entire content of a given Title,” and does not control for title size, id. at 363—are all nonpositive law titles, while nine of the 10 least complex titles are positive law titles. See id. at 366–67 tbl.11. Of course, Katz and Bommarito do not take statutory notes into account in their measure of complexity.

56. See Daniel B. Listwa, Uncovering the Codifier’s Canon: How Codification Informs Interpretation, 127 YALE L.J. 464, 475 (2017) (citing an email from Robert Sukol, Deputy Law Revision Counsel, explaining that the OLRC decides where to place provisions in nonpositive law titles based upon their own reading of the statute).
has the power to “classify newly enacted provisions of law” in nonpositive law titles without direction from Congress. The differences in the OLRC’s power with respect to positive and nonpositive law titles explains why some statutory notes exist.

Positive Law Titles—OLRC’s Forced Hand

¶16 Failure to amend positive law titles directly is one reason why statutory notes exist in the United States Code. When the OLRC receives a public law from Congress, it looks to see whether provisions in the statute amend an existing positive law title of the United States Code. If so, the changes are made as directed. For example, a portion of the recently passed FIRST STEP Act dealing with the use of restraints on prisoners during pregnancy provides: “Chapter 317 of title 18, United States Code, is amended by inserting after section 4321 the following . . . .” This is a direct amendment by Congress because it specifies what changes are to be made to a positive law title in the United States Code. In this instance and others like it, the OLRC simply follows the instructions given by Congress.

¶17 Many laws, however, do not directly amend positive law titles. They create freestanding provisions with no direction as to where in the Code they should go. In such cases, the OLRC’s first task is to determine whether the provisions of an act are general and permanent in nature. Provisions that are not general and permanent are not included in the Code. If a provision is general and permanent, then the OLRC must decide—through a process called classification—where the provision will be placed in the Code. If the OLRC classifies a provision in a nonpositive law title, the provision can simply be inserted into that title. If, on the other hand, the OLRC decides the provision should go into a positive law title, then it lacks the authority to insert the new law directly into the Code, and it must add it as a note.

¶18 Thus, in some instances, entire acts with important substantive provisions are relegated to statutory notes. The Torture Victim Protection Act of 1991 is such an example; this statute established a civil cause of action against someone who subjects an individual to torture or extrajudicial killing “under actual or apparent

57. 2 U.S.C. § 285b(4) (one of the functions of the OLRC is “[t]o classify newly enacted provisions of law to their proper positions in the Code where the titles involved have not yet been enacted into positive law.”).
60. See Tress, supra note 11, at 139.
62. Although OLRC has discretion to include the provisions as statutory notes as well. See infra notes 90–114 and accompanying text.
63. See Lynch, supra note 10, at 78; Nicholas Triffin, Questions and Answers, 77 Law Libr. J. 182, 184 (1984) (“[T]itles that have been enacted into positive law can be changed (by addition of new sections and amendment or repeal of existing sections) only by Congress. Thus, if Congress enacts a provision that should (by virtue of the subject matter) be set out in a title that has been enacted into positive law, but the provision was not enacted as an amendment of that title, we will classify the provision as a note under the appropriate section of the title.”). In some instances, the OLRC adds these types of acts to an appendix. See McKinney, supra note 19. For purposes of this article, we treat statutory provisions that end up in notes or an appendix the same because both are separate from code section text.
authority, or color of law, of any foreign nation.”64 In the text of the enacted bill, Congress did not specify the particular United States Code title where it would be placed. The OLRC decided it fit best near 28 U.S.C. § 1350, “Alien’s action for tort.” Title 28, however, is a positive law title, and because Congress did not draft it into title 28, the OLRC was powerless to do so. Consequently, the Act in its entirety is found in the United States Code only as a note.65

¶ 19 There are a few potential explanations for Congress’s failure to amend a positive law title directly to include an entire act. First, Congress may simply have made an error and neglected to amend a positive law title directly. Charles Zinn, who was in charge of the codification process shortly after positive law codification began, indicated that it had been a struggle to educate Congress to draft bills to amend positive law titles directly.66 He feared that as positive law titles increased, bills would “be drafted in general terms without reference to the applicable Title of the United States Code.”67 Federal drafting manuals emphasize that drafters must be very familiar with the differences between positive and nonpositive law titles so as to draft correctly.68 Without further investigation, it is impossible to say how often lack of direct amendment occurs simply by mistake, but the current system certainly invites error when legislation must be drafted a certain way depending on the title to which it will belong in the Code.

¶ 20 Even more likely than mistake is the possibility that the frenetic nature of a legislative session makes it difficult to get it right all of the time.69 Because of the time pressures of drafting legislation, the OLRC is in a better position to classify the Code, but its hands are tied if an act drafted as a freestanding law has a natural place in a positive law title. Because these new acts do not have a predestined place in the Code, legislative drafters may not have the time to decide where the act should appear in the Code. As the Torture Victim Protection Act teaches us, however, failure to designate a place for a new act in a positive law title can result in an unnatural placement of important laws in statutory notes.

¶ 21 While entire acts are often the most noticeable omissions from statutory text, portions of acts frequently become statutory notes. One example that raised concerns for researchers in the 1980s was the Privacy Act of 1974.70 Sections 3 and

65. 28 U.S.C. § 1350 note (2018). Besides the fact that a statutory note is more difficult to find than a statutory section, other concerns can also arise based on the positioning of an act as a note to a particular subsection rather than as an independent section. See William J. Aceves, Correcting an Evident Error: A Plea to Revise Jesner v. Arab Bank, PLC, 107 Geo. L.J. ONLINE 63, 63 (2018) (criticizing Justice Kennedy’s characterization of the Torture Victim Protection Act as a cause of action under the Alien Tort Statute based upon the OLRC’s placement of the statutory note).
66. See Zinn, supra note 45, at 394 (“One of our problems is trying to educate other committees of Congress to draft their bills in terms of the titles that have so far been enacted into law. We have been fairly successful to date, but as we broaden the scope of our work and get more into the field of other committees, I am afraid we are going to have a situation comparable to that which existed after the enactment of the Revised Statutes.”).
67. Id.
69. See ‘Tress, supra note 11, at 151.
70. Triffin, supra note 63, at 184; Alice I. Youmans et al., Questions and Answers, 78 LAW LIBR. 585, 592 (1986).
4 of that Act directly amended title 5 of the *United States Code*, while the remaining seven sections did not and were added as notes to 5 U.S.C. § 552a. Researchers were particularly concerned with section 7 of the Act, which made it unlawful for a government agency to deny benefits provided by law to anyone who refused to disclose his or her social security number. This important provision, which one federal appellate court described as one of only two substantive provisions in the Act, was relegated to a statutory note because it did not directly amend title 5 as did other sections of the same Act.

Similarly, when Congress passed a law regulating armor piercing ammunition in 1986, the Act defined “armor piercing ammunition,” in part, as a “projectile or projectile core which may be used in a handgun.” The legislative history of this Act shows there was concern in the Senate that an earlier amendment of the bill without a definition of “handgun” would “provide another giant loophole for the criminals.” A definition of “handgun” was eventually added before final passage of the bill, but it did not directly amend title 18 of the *United States Code* and was, therefore, inserted as a note to 18 U.S.C. § 921. Researchers using the definition of “armor piercing ammunition” found in 18 U.S.C. § 921(a)(17)(B) should be familiar with the applicable definition of “handgun.” Yet the definition had to be relegated to a statutory note.

Two more recent examples can be seen in the “America Invents Act” and the “FIRST STEP Act.” Passed in 2011, the “America Invents Act” was a significant revision of patent law. During debate over the Act, Congressman Smith of Texas offered an amendment that added a provision prohibiting the patenting of inven-

---

72. Id. § 7, 88 Stat. at 1909.
73. Schwier v. Cox, 340 F.3d 1284, 1287 (11th Cir. 2003).
74. Interestingly, in this example one commentary opined that the inclusion of § 7 as a note “helps to [e]nsure that researchers will not overlook that portion of the statute.” Youmans et al., supra note 70, at 592. While including the section in a note is surely a better option than not including it at all, it begs the question of how it is possible that an important provision is not included in a statutory section.
76. 132 Cong. Rec. 3893 (Mar. 6, 1986).
77. The last section of the Act states:
   For purposes of section 921(a)(17)(B) of title 18, United States Code, as added by the first section of the Act, “handgun” means any firearm including a pistol or revolver designed to be fired by the use of a single hand. The term also includes any combination of parts from which a handgun can be assembled.
   § 10, 100 Stat. at 922.
78. An interesting question arises here. The Brady Handgun Violence Prevention Act passed in 1993 added its own definition of “handgun” to 18 U.S.C. § 921. While the definitions of “handgun” in 18 U.S.C. § 921(a)(29) and 18 U.S.C. § 921 note are similar, they are not identical. Researchers looking today at the definition of “armor piercing ammunition” under 18 U.S.C. § 921(a)(17) are likely to scan the rest of the definitions and come across the definition of 18 U.S.C. § 921(a)(29) and think that it’s applicable. 18 U.S.C. § 921 note, however, tells us explicitly that for purposes of 18 U.S.C. § 921(a)(17) its definition of handgun applies. On the other hand, 18 U.S.C. § 921(a) tells us that the definitions apply to the chapter.
tions “directed to or encompassing a human organism.”81 While most of the substantive changes to title 35, a positive law title, directly amended current statutory sections, this amendment did not. The amendment was agreed to and became part of the final bill, leaving this provision as a statutory note to 35 U.S.C. § 101.

¶24 As with the “America Invents Act,” the bulk of the “FIRST STEP Act” directly amended sections of title 18 of the United States Code, properly placing the provisions in code sections. Some provisions of this criminal justice reform legislation, however, including provisions regarding reports, appropriations, statutory construction, faith-based considerations, data collection, and a provision requiring the Director of the Bureau of Prisons to make feminine hygiene products available to prisoners for free, did not explicitly amend title 18.82 Because title 18 is a positive law title, OLRC’s hands are tied and options limited: these provisions can only be added to title 18 as statutory notes.83

¶25 In each of these examples, it is clear that Congress knew that the title in question needed to be directly amended because at least some of the act did so. What is less certain, then, is why certain provisions were not directly amended. One possibility, as mentioned above, is that the hustle and bustle of the legislative process can produce such an oversight. This seems to have been the case with the armor piercing ammunition law that added a definition for “handgun” toward the end of the legislative process to appease some senators. The original bill properly amended title 18, but instead of incorporating the new definition of “handgun” into the amending portions of the bill, the Senate simply tacked it onto the end of the bill, banishing it—likely unwittingly—to a statutory note.

¶26 Another possibility, which likely occurs more frequently, has to do with preferences in legislative drafting. Frequently, statutory provisions that do not directly amend positive law titles are of a certain type, such as short titles, effective dates, appropriations, severability provisions, rules of construction, and directions to agencies. At least one federal legislative drafter points out that nonbinding provisions are excluded from the Code.84 Other legislative drafting manuals show a distaste among drafters for including provisions in legislation that are “not legally useful,” such as “findings that are nothing more than rhetoric, definitions that merely state the obvious, and precatory language . . . that has no binding effect.”85 Drafters are also encouraged to avoid severability provisions because severability is the default rule.86 When these types of provisions are included in legislation that is meant for a positive law title, it is likely that legislative drafters are purposefully avoiding direct amendment so as not to include these types of provisions in the

82. See FIRST STEP Act, supra note 79.
83. The other option OLRC has is to find a nonpositive law title in which to include these provisions or leave them out of the Code completely.
84. Ryneason, supra note 68, at 83.
85. See Dorsey, supra note 58, at 177; see also Filson & Strokov, supra note 68, at 1217 (discouraging findings and purposes because they are not legally useful).
86. Senate Legislative Drafting Manual 49 (“Consequently, a severability clause is unnecessary.”); House Drafting Manual 32–33 (explaining that general severability provisions are unnecessary); Filson & Strokov, supra note 68, at 182–83 (calling inclusion of severability clauses in legislation a “questionable practice”); Dorsey, supra note 58, at 228; Donald Hirsch, Drafting Federal Law 20 (1980).
code. The sections of the Privacy Act of 1974, for example, that did not directly amend title 5 included a short title, findings, creation of the Privacy Protection Study Commission, directions to the Office of Management and Budget, an effective date, and appropriations. The FIRST STEP Act reveals similar provisions that likely purposefully did not directly amend title 18.

27 To be sure, when a positive law title is not amended directly—that is, when the legislation fails to specify the title and section being amended—the result is the same: the act or provisions must be excluded from the positive law title because the OLRC does not have the authority to insert a new law directly into the Code. Thus, those provisions are added as statutory notes. Being familiar with this rule can help researchers working in areas covered by positive law titles be more aware of the possibility that substantive (and other relevant) provisions of law could be found in statutory notes.

Nonpositive Law Titles

28 While failure to amend positive law titles directly may seem like a strange reason for the existence of statutory notes, at least the rule is clear: if Congress does not directly amend a Code section and the OLRC classifies it in a positive law title, it must appear as a statutory note. In certain instances, however, failure to amend a positive law title directly is not the cause of a statutory note. Statutory notes appear in nonpositive law titles as well. In fact, the first statutory notes appeared in the 1940 edition of the United States Code before any title was enacted into positive law. The preface to that edition lists “additional notes” as one of the reasons the Code was expanded to four volumes. Examination of the 1940 edition reveals a limited number of statutory notes, including savings clause provisions, effective date provisions, and some more substantive provisions. The inclusion of statutory notes in nonpositive law titles in 1940 and today is an editorial decision made by the OLRC. Unfortunately, because of the complex nature of the United States Code and the classification process, there is no single, simple explanation for when statutory provisions become statutory notes in nonpositive law titles.

87. A potential downside of positive law codification is that it will likely increase the number of statutory notes in the United States Code because legislative drafter will avoid direct amendment in order to keep out provisions of these types that the OLRC has shown a willingness to include in the Code.

88. Privacy Act of 1974, Pub. L. No. 93-579, 88 Stat. 1896. One section that did not directly amend title 5 and that seems to be a substantive provision was section 7, which forbid federal, state, or local agencies from denying individuals any right, benefit, or privilege provided by law based on a failure to provide a Social Security Number. Id. at 1909.

89. See supra note 82 and accompanying text.

90. See generally United States Code (1940).

91. Preface of United States Code, at ix (1940) (other stated reasons included the addition of a number of laws and the use of “a more legible type”).


94. E.g., 42 U.S.C. § 1001 note (1940) (providing that tax not be collected under certain acts for services, rendered prior to January 1, 1940, for salvaging timber or clearing debris left by hurricane).

¶29 The OLRC’s general preference is to classify statutory provisions as Code sections rather than statutory notes.⁹⁶ Even with that general preference, however, there are several types of provisions that the OLRC “normally” classifies as statutory notes.⁹⁷ These include “effective dates, short titles, savings, and statutory construction,” as well as findings and severability provisions.⁹⁸ Additionally, “provisions that are somewhat less than general or permanent, but still relate to existing Code sections, such as those requiring studies or reports, implementation of regulations, or the establishment of a task force” normally become statutory notes.⁹⁹ Researchers should not be surprised to find these types of provisions appear as statutory notes in nonpositive law titles.

¶30 These exceptions to the general preference of classifying provisions as Code sections have exceptions of their own, however. For example, provisions normally classified as statutory notes “should be classified as sections where they are part of an entire hierarchical unit that is being added to the Code.”¹⁰⁰ Additionally, “[p]rovisions involving the creation of an office should be classified as sections, while provisions that establish minor offices should be classified as notes.”¹⁰¹ Even more general provisions that might otherwise be classified as a code section may be classified as notes if they are “closely tied to an existing Code section.”¹⁰²

¶31 One of the principles that emerges for why some statutory provisions end up as statutory notes in nonpositive law titles is act-Code coherence. This is the idea that, to the extent possible, the OLRC attempts to keep the language of an act together “in order to make the law accessible and understandable to the reader.”¹⁰³ Act-Code coherence does not explain all statutory notes, but it can help researchers understand certain decisions that at first may seem confusing. Take, for example, a provision from the Immigration Technical Corrections Act of 1988 that clarified a definition found in 8 U.S.C. § 1101(a)(27)(I), regarding residence in the United States of an employee of an international organization. Because title 8 is a nonpositive law title, the OLRC could conceivably have included this as a separate subsection of 8 U.S.C. § 1101(a)(27)(I). However, this provision did not amend the underlying Immigration and Nationality Act (INA), and simply inserting it would have broken up the INA as it appears in the Code.¹⁰⁴ Instead, this provision was added as

---

⁹⁶ See Email from Ralph Seep, Law Revision Counsel, U.S. House of Representatives, to authors (June 25, 2019, 09:53 EST) (on file with authors).
⁹⁷ About Classification, supra note 61.
⁹⁸ Id.; accord Email from Ralph Seep, supra note 96.
⁹⁹ About Classification, supra note 61.
¹⁰⁰ Email from Ralph Seep, supra note 96.
¹⁰¹ Id.
¹⁰² Id.
¹⁰³ Id.; see also McKinney, supra note 19 (“Generally, it seems, the codifiers like to place a statute’s language all together in the U.S. Code without breaks from other statutes.”). Detailed Guide, supra note 1 (“In the case of a nonpositive law title, the organization of the title since 1926 has been determined by the editors of the Code and has generally followed the organization of the underlying acts as much as possible.”). Often the complexity of federal legislation that spans multiple topics makes it impossible to maintain act-Code coherence. See About Classification, supra note 61.
¹⁰⁴ In order to force the provision into the INA, Congress would have had to amend the INA by stating something like, “Section x of the Immigration and Nationality Act is amended to read . . . .” In the absence of such a directive, the OLRC would generally avoid placing the new provision in statutory text in the midst of the INA. See McKinney, supra note 19 (stating that it has been “a fairly hard rule in recent decades” not to “combine a new statutory provision with an existing code section
a statutory note, despite being essential to understanding 8 U.S.C. § 1101(a)(27)(I). In fact, this provision is so essential that when the OLRC prepared a bill to enact title 8 into positive law in 1995, it moved this provision out of the notes and into a section.\textsuperscript{105} The bill ultimately failed, leaving title 8 as a nonpositive law title and this definition as a note.

\textsection{32} Similarly, an important statutory provision affecting the Fair Labor Standards Act is only included as a statutory note in the \textit{United States Code}. Section 207 of the Fair Labor Standards Act, found at 29 U.S.C. § 207, requires employers to pay overtime to employees working more than 40 hours a week.\textsuperscript{106} However, 29 U.S.C. § 213(b)(1) exempts certain employees from overtime compensation.\textsuperscript{107} In 2008, Congress passed a law that removed a class of employees from the 213(b)(1) exemption.\textsuperscript{108} Unfortunately, Congress did not amend the FLSA when passing this provision, so OLRC did not insert it as a code section or part of a code section. Instead, it did the next best thing and inserted that provision and its definition of “covered employee” as a statutory note.

\textsection{33} In other instances, the OLRC attempts to be faithful to the structure of an act passed by Congress even if this means relegating certain provisions to statutory notes. Congress passed the Better Online Ticket Sales Act of 2016 (BOTS Act) to “prohibit the circumvention of control measures used by Internet ticket sellers.”\textsuperscript{109} Section 2 of the Act, which enumerates the prohibited conduct, exceptions, and enforcement of the Act, was added as a new statutory section in title 15.\textsuperscript{110} Section 3 of the Act, however, which provides the essential definitions for section 2—including “commission,” “event,” “event ticket,” and “ticket issuer”—was added as a statutory note.\textsuperscript{111}

\textsection{34} To understand the placement of these definitions in a statutory note, it’s useful to compare the BOTS Act with the Consumer Review Fairness Act (CRFA). Both were passed on December 14, 2016, and ended up beside each other in the \textit{United States Code} as §§ 45b and 45c of title 15.\textsuperscript{112} The Acts were drafted slightly differently, however, leading to a major difference in the way their definitions are presented in the Code. As in the BOTS Act, section 2 of the CRFA provides the heart of the Act and is included as a Code section. Unlike the BOTS Act, however, representing a previous Act that Congress did not specifically amend”).

\textsuperscript{105} H.R. 1292, 104th Cong. (1995).
\textsuperscript{106} 29 U.S.C. § 207.
\textsuperscript{107} 29 U.S.C. § 213(b)(1).
\textsuperscript{111} See 15 U.S.C. § 45c note. In addition to highlighting the importance of examining the statutory notes, this example also demonstrates the importance of editorial notes. A researcher analyzing 15 U.S.C. § 45c may look for definitions nearby in the table of contents. Doing so would lead them to 15 U.S.C. § 44, which provides definitions for the subchapter. However, a codification note for 15 U.S.C. § 45c provides that it was enacted as part of the BOTS Act of 2016 “and not as part of the Federal Trade Commission Act which comprises this subchapter.” So, while § 45c is included in the Federal Trade Commission Act subchapter, its inclusion as part of the subchapter was an editorial decision.
\textsuperscript{112} Compare Consumer Review Fairness Act of 2016, Pub. L. No. 114-258, 130 Stat. 1355, with BOTS Act of 2016, Pub. L. No. 114-274, 130 Stat. 1401. It is useful to point out that the addition of these new sections would not have been possible if title 15 were positive law.
the definitions of the CRFA are included in section 2 and are, therefore, part of the Code section. In the BOTS Act, the definitions were included in a final, separate section that begins “In this Act.” In order to be faithful to the structure of the Act, the OLRC did not include this definitions section as a separate statutory section. While it is understandable that the OLRC would be deferential to acts as passed by Congress, from a researcher’s perspective it is troubling to have two neighboring sections laid out so differently.

Statutory notes exist in nonpositive law titles for a variety of reasons. OLRC’s general preference is to classify statutes as code sections rather than statutory notes, but classification is extremely complicated and has many moving parts. OLRC attorneys are guided by a number of rules and principles when deciding whether to create a statutory note, but the decision is ultimately editorial. Where possible, OLRC tries to maintain act-Code coherence or to otherwise be faithful to the structure as passed by Congress, which leads to some similar provisions being placed in code sections and others in statutory notes. Because there is no definitive rule on when certain provisions end up in statutory notes in nonpositive law titles, researchers must always be on the lookout for these provisions.

Extent of the Problem

Earlier research has identified the problem of statutory notes in the United States Code but has stopped short of attempting to determine the extent of the problem. A likely reason this research has not been attempted is its difficulty. As discussed above, statutory notes are mixed with a variety of other notes, and there is no easy way to separate relevant statutory notes from others. One way to get some idea of the number of laws that are relegated to statutory notes is to examine the number and types of notes that appear in court opinions. While this will not uncover all substantive or legally significant provisions found in statutory notes, it can provide a good picture of how often these statutory notes are used by litigants and judges and provide examples of important laws that are found in statutory notes. To limit the results to a manageable number, we searched U.S. Supreme

113. A companion bill, H.R. 5104 (2016), incorporated its definitions into the heart of the Act. If this bill would have passed, it is likely that the definitions would be part of 45b, not a statutory note. See H.R. 5104, 114th Cong. (2016).

114. While the CRFA and the BOTS Act are examples of OLRC being faithful to the structure Congress gave to the Acts, they are also examples of OLRC bending the rule of act-Code coherence. OLRC inserted both acts in the middle of a subchapter of title 15 that contains the Federal Trade Commission Act. This insertion is mentioned in an editorial note after each section. 15 U.S.C. § 45b note; 15 U.S.C. § 45c note.

115. Michael Lynch remarked that “the way seems tedious” in pursuing a project like this. Lynch, supra note 10, at 80.

116. To determine the number of cases that referred to notes, the authors began by searching all state and federal cases in Westlaw for note /5 U.S.C. This brought back 10,000 results, which is Westlaw’s maximum, meaning that there were more than 10,000 cases corresponding to this result. After examining some of these results, it became clear that “44 U.S.C. s 3501 note” would need to be excluded from the search because this statutory note contains the E-Government Act of 2002, which requires decisions of the United States Court of Federal Claims to be posted to its website. This instruction, with a citation to 44 U.S.C. § 3501 note, appears in the first footnote of many cases from the United States Court of Federal Claims since 2002. See, e.g., Encinias v. Sec’y, Health & Human Servs., 2017 WL 2417773 (Fed. Cl. 2017) (It appears that there are nearly 8,000 cases that contain a citation to “44 U.S.C. s. 3501 note.” This is interesting in itself because many of these are about disputes regarding a litigant’s right to prevent certain decisions from being posted to the Federal Court of
Court and U.S. court of appeals decisions in Westlaw from 2008 to 2018. After eliminating false positives, we found 605 U.S. Supreme Court and U.S. court of appeals cases over this 10-year period that cited a statutory note. Forty-eight of these cases came from the U.S. Supreme Court. We also examined a smaller slice

Claims website). With this amendment, the most fruitful search for finding statutory notes in caselaw was (note /5 U.S.C.) & DA(aft 12-31-2007 & bef 01-01-2019) % "44 U.S.C. s 3501". This search still resulted in over 10,000 results, so the authors determined that it would be beneficial to restrict the results to the last 10 years. Doing so whittled the results down to 4854 cases. While this was more manageable, the authors decided to focus on U.S. Supreme Court and Federal Court of Appeals opinions to bring the number of opinions down even more.

117. The exact search used was (note /5 U.S.C.) & DA(aft 12-31-2007 & bef 01-01-2019) % "44 U.S.C. s 3501". This resulted in 1316 cases before irrelevant results were removed.


119. It's important to point out that this number is likely underinclusive because of the way statutory notes are cited in court opinions. In 1997, Michael Lynch observed that researchers citing to a statutory note must cite to the Statutes at Large. See Lynch, supra note 10, at 79. The Bluebook seemed to support this approach until 2015, when the 20th edition added a section directing that references to statutory notes cite to the United States Code. See The Bluebook: A Uniform System of Citation R. 12.3.1(h), at 124 (Columbia Law Review Ass'n et al. eds., 20th ed. 2015). In practice, some courts follow The Bluebook, while others cite the Statutes at Large and the United States Code as parallel citations. There seems to be little consistency in how statutory notes are cited. Because of this inconsistency, it is likely that some opinions have cited only to Statutes at Large and are missed by the above search. See, e.g., Youmans et al., supra note 70, at 590–91 (citing a district court decision, Greater Cleveland Welfare Rights Org. v. Bauer, 462 F. Supp. 1313 (N.D. Ohio 1978), that cited to the public law version of the Privacy Act of 1974 rather than its United States Code note citation); Glenn v. Holder, 690 F.3d 417, 420 (6th Cir. 2012) (citing § 4710 of the Hate Crimes Act instead of 18 U.S.C. § 249 note).

of U.S. district court cases and found 514 cases citing statutory notes over the five-year period from 2014 to 2018.

¶37 Examining these cases more closely reveals that federal courts cite a wide variety of notes. Some of the more commonly cited acts that appear as statutory notes include the Torture Victim Protection Act,\textsuperscript{121} the Terrorism Risk Insurance Act,\textsuperscript{122} the General Aviation Revitalization Act,\textsuperscript{124} the Internet Tax Freedom Act,\textsuperscript{125} and the Hyde Amendment.\textsuperscript{126} Other statutory notes cited in case law include directions to agencies,\textsuperscript{127} rules of construction,\textsuperscript{128} effective dates,\textsuperscript{129} findings and purposes,\textsuperscript{130} and savings provisions.\textsuperscript{131} Some notes even appear to be provisions that the OLRC may have considered “less than general or less than permanent” enough to become code sections, including many in the area of immigration and nationality law, such as the Nicaraguan Adjustment and Central American Relief Act\textsuperscript{132} and the Cuban Refugee Adjustment Act of 1966.\textsuperscript{133}

¶38 A closer look at these results reveals an interesting comparison between statutory notes in positive law titles and those in nonpositive law titles. In federal appellate courts, statutory notes in positive law titles were cited in 373 cases, while

---

\textsuperscript{121} 28 U.S.C. § 1350 note.
\textsuperscript{122} 28 U.S.C. § 1610 note.
\textsuperscript{123} 46 U.S.C. § 30701 note.
\textsuperscript{124} 49 U.S.C. § 40101 note.
\textsuperscript{125} 47 U.S.C. § 151 note.
\textsuperscript{126} 18 U.S.C. § 3006A note.
\textsuperscript{129} Goodwin v. Reynolds, 757 F.3d 1216, 1218 n.1 (11th Cir. 2014) (citing the effective date of a 2011 amendment to 28 U.S.C. § 1441 note).
those in nonpositive law titles were cited in 246 cases.\textsuperscript{134} In federal district courts, 345 statutory note citations were to positive law titles while 172 were to nonpositive law titles. At the outset of our research, we had guessed that because of the nature of the amendatory process for positive law titles, statutory notes in these titles would have appeared much more frequently in case law than those in nonpositive law titles. While statutory notes from positive law titles were more common, the disparity was not as great as we expected. This important finding demonstrates that researchers cannot focus merely on notes in positive law titles, but must be on the lookout for statutory notes in nonpositive law titles as well.\textsuperscript{135}

¶39 While surveying case law will provide some indication of the extent to which statutory notes are used in practice, citations tell only a fraction of the story. Likely, only a small portion of statutory notes will ever appear in court opinions. In order to get an accurate count of statutory notes, we identified textual patterns common to all statutory notes. While statutory notes cannot be identified by a particular title, they generally begin with the phrase “Pub. L. [public law number] provided that:” followed by a quotation of statutory text.\textsuperscript{136} After identifying these patterns, we ran a Python script on the 2018 edition of the \textit{United States Code} to extract all statutory notes and, after excluding short title and effective date statutory notes, found 14,522 statutory notes in the 2018 \textit{United States Code}. Including short title and effective date statutory notes raises the total to 32,424.\textsuperscript{137} Table 1 shows the breakdown of statutory notes by title of the \textit{United States Code}, organized by the number of statutory notes, exclusive of short title and effective dates. The total number of statutory notes is also included, along with a size rank of each title so that researchers can compare the relative size to the number of statutory notes.\textsuperscript{138} Positive law titles are marked with an asterisk.

¶40 This title breakdown can be helpful to researchers as they navigate different portions of the \textit{United States Code}. Researchers working in title 10 (Armed Forces) or title 49 (Transportation), for example, are much more likely to encounter statutory notes than those focused on title 11 (Bankruptcy). This table provides additional evidence that statutory notes are just as prominent in nonpositive law titles as positive law titles. In fact, more than half of statutory notes are found in nonpositive law titles.

\footnotesize{134. This total exceeds the number of cases citing statutory notes because some cases cite multiple statutory notes. See, e.g., Vance v. Rumsfeld, 701 F.3d 193, 198, 201, 206 (7th Cir. 2012) (citing the Detainee Treatment Act, 10 U.S.C. §801 note, and the Torture Victim Protection Act, 28 U.S.C. §1350 note).

135. Statutory notes in title 8, Aliens and Nationality, made up nearly 40 percent of citations of nonpositive law notes in federal appellate cases and nearly 15 percent in federal district court cases. It is likely that statutory notes are prevalent in title 8 because the statutory provisions target groups of people and may not be considered generally applicable. Future researchers may want to explore further what impact, if any, this has on immigration law and policy.

136. Other versions include, “Act . . . provided that:” and “Section . . . provided that:.”

137. At the beginning of our research, we estimated the number of statutory notes based on a search of the \textit{United States Code} index. Searching for statutory notes in the 2012 \textit{United States Code} index results in 44,491 entries for statutory notes, which we believed to be overinflated because index terms are indexed under more than one heading. The total of 32,424 statutory notes appears to be a plausible number based on the number of index entries.

138. The size rank was first used by Katz and Bommarito in 2010 and is based on the number of “tokens” in each title of the Code. “Tokens” are described as “contiguous strings of text, which are often words but may also be numbers, citations, or abbreviations not formally considered words.” Katz & Bommarito, supra note 55, at 353. We used the Katz and Bommarito definition to update the rankings based on the 2018 U.S. Code.}
<table>
<thead>
<tr>
<th>Title</th>
<th>Statutory Notes (Excluding Short Titles &amp; Eff. Dates)</th>
<th>Total Statutory Notes</th>
<th>Size of Title Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>42-The Public Health and Welfare</td>
<td>2280</td>
<td>5737</td>
<td>1</td>
</tr>
<tr>
<td>*10-Armed Forces</td>
<td>1922</td>
<td>2802</td>
<td>7</td>
</tr>
<tr>
<td>26-Internal Revenue Code</td>
<td>964</td>
<td>5909</td>
<td>2</td>
</tr>
<tr>
<td>*49-Transportation</td>
<td>844</td>
<td>1024</td>
<td>10</td>
</tr>
<tr>
<td>22-Foreign Relations and Intercourse</td>
<td>705</td>
<td>1158</td>
<td>8</td>
</tr>
<tr>
<td>15-Commerce and Trade</td>
<td>585</td>
<td>1242</td>
<td>4</td>
</tr>
<tr>
<td>16-Conservation</td>
<td>552</td>
<td>1036</td>
<td>3</td>
</tr>
<tr>
<td>*5-Government Organization and Employees</td>
<td>498</td>
<td>1036</td>
<td>17</td>
</tr>
<tr>
<td>*38-Veterans' Benefits</td>
<td>489</td>
<td>1317</td>
<td>13</td>
</tr>
<tr>
<td>7-Agriculture</td>
<td>445</td>
<td>888</td>
<td>5</td>
</tr>
<tr>
<td>12-Banks and Banking</td>
<td>388</td>
<td>839</td>
<td>6</td>
</tr>
<tr>
<td>21-Food and Drugs</td>
<td>362</td>
<td>654</td>
<td>11</td>
</tr>
<tr>
<td>*28-Judiciary and Judicial Procedure</td>
<td>327</td>
<td>562</td>
<td>27</td>
</tr>
<tr>
<td>50-War and National Defense</td>
<td>276</td>
<td>450</td>
<td>18</td>
</tr>
<tr>
<td>8-Aliens and Nationality</td>
<td>274</td>
<td>571</td>
<td>24</td>
</tr>
<tr>
<td>*23-Highways</td>
<td>271</td>
<td>375</td>
<td>31</td>
</tr>
<tr>
<td>*31-Money and Finance</td>
<td>269</td>
<td>394</td>
<td>29</td>
</tr>
<tr>
<td>33-Navigation and Navigable Waters</td>
<td>238</td>
<td>395</td>
<td>15</td>
</tr>
<tr>
<td>*18-Crimes and Criminal Procedure</td>
<td>229</td>
<td>571</td>
<td>16</td>
</tr>
<tr>
<td>2-The Congress</td>
<td>217</td>
<td>481</td>
<td>22</td>
</tr>
<tr>
<td>19-Customs Duties</td>
<td>211</td>
<td>478</td>
<td>14</td>
</tr>
<tr>
<td>20-Education</td>
<td>183</td>
<td>583</td>
<td>9</td>
</tr>
<tr>
<td>*46-Shipping</td>
<td>182</td>
<td>245</td>
<td>25</td>
</tr>
<tr>
<td>43-Public Lands</td>
<td>163</td>
<td>233</td>
<td>20</td>
</tr>
<tr>
<td>47-Telecommunications</td>
<td>141</td>
<td>257</td>
<td>26</td>
</tr>
<tr>
<td>29-Labor</td>
<td>128</td>
<td>376</td>
<td>12</td>
</tr>
<tr>
<td>*37-Pay and Allowances of the Uniformed Services</td>
<td>114</td>
<td>380</td>
<td>37</td>
</tr>
<tr>
<td>25-Indians</td>
<td>102</td>
<td>223</td>
<td>19</td>
</tr>
<tr>
<td>6-Domestic Security</td>
<td>96</td>
<td>152</td>
<td>28</td>
</tr>
<tr>
<td>*51-National and Commercial Space Programs</td>
<td>85</td>
<td>103</td>
<td>46</td>
</tr>
<tr>
<td>*14-Coast Guard</td>
<td>81</td>
<td>103</td>
<td>41</td>
</tr>
<tr>
<td>34-Crime Control and Law Enforcement</td>
<td>76</td>
<td>246</td>
<td>21</td>
</tr>
<tr>
<td>*44-Public Printing and Documents</td>
<td>68</td>
<td>114</td>
<td>43</td>
</tr>
<tr>
<td>*41-Public Contracts</td>
<td>64</td>
<td>87</td>
<td>38</td>
</tr>
<tr>
<td>30-Mineral Lands and Mining</td>
<td>59</td>
<td>123</td>
<td>23</td>
</tr>
<tr>
<td>*1-General Provisions</td>
<td>58</td>
<td>62</td>
<td>56</td>
</tr>
</tbody>
</table>
tive law titles, although this is partially because nonpositive law titles make up approximately three-quarters of the *United States Code*. It is important to note, however, that for many positive law titles, the number of statutory notes is larger than would be expected based on the size of the particular title. While this is true for some nonpositive law titles, it appears to be more true for positive law titles and gives credence to the idea that positive law titles increase the possibility for statutory notes because OLRC cannot add provisions to code sections themselves if provisions do not directly amend a positive law title.

¶ 41 An additional way to understand the extent of statutory notes is to examine the OLRC’s classification table, known as Table III. Table III provides information about where each section of a public law was classified in the *United States Code*. Table III is helpful for researchers who want to track where provisions of a public

<table>
<thead>
<tr>
<th><em>United States Code</em> title</th>
<th>Statutory Notes (Excluding Short Titles &amp; Eff. Dates)</th>
<th>Total Statutory Notes</th>
<th>Size of Title Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>39-Postal Service</em></td>
<td>58</td>
<td>125</td>
<td>44</td>
</tr>
<tr>
<td>48-Territories and Insular Possessions</td>
<td>58</td>
<td>97</td>
<td>34</td>
</tr>
<tr>
<td>5-Appendix</td>
<td>57</td>
<td>95</td>
<td>48</td>
</tr>
<tr>
<td><em>40-Public Buildings, Property, and Works</em></td>
<td>56</td>
<td>75</td>
<td>35</td>
</tr>
<tr>
<td><em>17-Copyrights</em></td>
<td>51</td>
<td>134</td>
<td>33</td>
</tr>
<tr>
<td>45-Railroads</td>
<td>46</td>
<td>141</td>
<td>30</td>
</tr>
<tr>
<td><em>35-Patents</em></td>
<td>43</td>
<td>130</td>
<td>47</td>
</tr>
<tr>
<td><em>36-Patriotic and National Observances, Ceremonies, and Organizations</em></td>
<td>40</td>
<td>51</td>
<td>36</td>
</tr>
<tr>
<td><em>54-National Park Service and Related Programs</em></td>
<td>34</td>
<td>61</td>
<td>42</td>
</tr>
<tr>
<td><em>13-Census</em></td>
<td>26</td>
<td>37</td>
<td>53</td>
</tr>
<tr>
<td><em>11-Bankruptcy</em></td>
<td>25</td>
<td>83</td>
<td>32</td>
</tr>
<tr>
<td><em>32-National Guard</em></td>
<td>23</td>
<td>32</td>
<td>50</td>
</tr>
<tr>
<td>52-Voting and Elections</td>
<td>17</td>
<td>67</td>
<td>40</td>
</tr>
<tr>
<td><em>3-The President</em></td>
<td>15</td>
<td>24</td>
<td>52</td>
</tr>
<tr>
<td>24-Hospitals and Asylums</td>
<td>9</td>
<td>16</td>
<td>51</td>
</tr>
<tr>
<td><em>4-Flag and Seal, Seat of Government, and the States</em></td>
<td>8</td>
<td>19</td>
<td>55</td>
</tr>
<tr>
<td>18-Appendix</td>
<td>5</td>
<td>12</td>
<td>49</td>
</tr>
<tr>
<td>11-Appendix</td>
<td>2</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td>27-Intoxicating Liquors</td>
<td>2</td>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td>28-Appendix</td>
<td>1</td>
<td>8</td>
<td>39</td>
</tr>
<tr>
<td><em>9-Arbitration</em></td>
<td>0</td>
<td>2</td>
<td>57</td>
</tr>
</tbody>
</table>

139. Of 14,522 statutory notes (excluding short titles and effective dates), 8,642 are found in nonpositive law titles.

law have gone once added to the Code, which also makes it a useful tool in understanding how many provisions end up as statutory notes. Because Table III provides information about statutory notes that have been repealed or eliminated, it is difficult to arrive at an accurate count of current notes that exist in the Code through a simple search.\(^\text{141}\) We can look closely at certain years, excluding notes that have been repealed or eliminated, to get a snapshot of the number of statutory notes entering the Code during a particular time period. Over the five-year period, seen in Table 2, from 2012 to 2016, 3938 statutory notes were added to the United States Code. These notes came from 522 unique public laws that were classified to the United States Code, meaning that during this time period, each public law produced, on average, approximately 7.5 statutory notes. These numbers include short titles and effective dates, but comparing these numbers to the overall statutory note total of 32,424, we find that 12% of all statutory notes were added during this 5-year time period. It appears likely this trend will continue.

### Table 2
Statutory Notes, 2012–2016

<table>
<thead>
<tr>
<th>Year</th>
<th># of Public Laws classified</th>
<th># of statutory notes</th>
<th>Average notes per Public Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>95</td>
<td>554</td>
<td>5.8</td>
</tr>
<tr>
<td>2013</td>
<td>84</td>
<td>674</td>
<td>8.0</td>
</tr>
<tr>
<td>2014</td>
<td>127</td>
<td>917</td>
<td>7.2</td>
</tr>
<tr>
<td>2015</td>
<td>81</td>
<td>971</td>
<td>12.0</td>
</tr>
<tr>
<td>2016</td>
<td>135</td>
<td>822</td>
<td>6.1</td>
</tr>
<tr>
<td>Total</td>
<td>522</td>
<td>3,938</td>
<td>7.5</td>
</tr>
</tbody>
</table>

\(^\text{141}\) A search for \textit{nt} in Table III results in 62,061 hits. The way Table III is set up makes it difficult to perform a simple search to eliminate false hits that occur when the status of the note is labeled as “Elim.” (eliminated) or “Rep.” (repealed).

\(^\text{142}\) See \textit{About Classification}, supra note 61; \textit{Detailed Guide}, supra note 1.

\(^\text{143}\) See Sanders v. Allison Engine Co., Inc., 703 F.3d 930, 939 n.6 (6th Cir. 2012) (“Section 4(f) of FERA is not codified in the text of § 3729, and is instead in the historical and statutory notes accompanying the section . . . The United States argues that this counsels against crediting the argument that the definitions in § 3729, which specifically apply for “purposes of this section,” should be used to define terms used in § 4(f) or in the note to § 3729.”); Schwier v. Cox, supra note 10, at

---

\[\text{The Shadow Code}\]

\[\text{42}\] The structure and organization of the Code itself, with its division between code sections and statutory notes, fosters the idea that statutory notes are something less than law. This idea is so pervasive that OLRC points out multiple times on its website that the placement of a statutory provision in a statutory note does not affect its validity as law.\(^\text{142}\) Court opinions show that federal judges and attorneys have at times misrepresented the significance of statutory notes and how they operate.\(^\text{143}\) Law librarians have even questioned whether statutory notes are valid
law.\textsuperscript{144} This confusion is understandable because the division between laws in code sections and laws in statutory notes is not intuitive. The official print version of the United States Code even puts statutory notes in a smaller font, further obscuring their importance.\textsuperscript{145}

\textsection{43} The Code’s structure casts a shadow over valid laws that exist in statutory notes, making them far more difficult to find. For example, a researcher could read the text of 29 U.S.C. \textsection{213}, the Fair Labor Standards Act provision referenced above that exempts certain employees from overtime compensation under 29 U.S.C. \textsection{207}, and have no idea that there is a statutory provision that removes a certain class of employees from this exemption. Without knowing better, this researcher would have no reason to stray beyond the text of the statute she has read. Fortunately, legal researchers can turn to other resources such as secondary sources, cases, and administrative materials to understand that this provision exists, but should this be necessary? Novice researchers would have little chance to discover this and other important provisions.

\textsection{44} While the simple fact that some binding law appears in statutory notes at all is enough to obscure these provisions, several other factors push statutory notes further into the shadows. First, the classification process leads to the placement of similar provisions in both Code sections and statutory notes. This lack of consistency makes it difficult for researchers to know when they should look for statutory notes. Second, the placement of editorial and statutory notes together with no clear distinction makes it difficult for researchers to find statutory notes or to understand their importance. Finally, legal research systems, primarily Westlaw, hide statutory notes from researchers, making them nearly invisible unless a researcher knows where to look.

\begin{itemize}
\item 1288; Pennie v. Twitter, Inc., 281 F. Supp. 3d 874, 889 (N.D. Cal. 2017) (inferring that codification of statutory note necessary to amend another statutory provision); Sciortino v. PepsiCo, Inc., 108 F. Supp. 3d 780, 797 n.5 (N.D. Cal. 2015) (court rejected defendant’s position at oral argument that statutory note from Nutrition Labeling and Education Act should not merit weight); Murkledove v. Astrue, 635 F. Supp. 2d 564, 580 (N.D. Tex. 2009) (“To begin with, the 1985 enactment of a note cannot be used to change the clear language of the 1980 amendment of the EAJA.”); Baez v. United States, 715 F. Supp. 2d 1165, 1178 (D. Or. 2010) (court disagreed with defendant that Cuban Refugee Adjustment Act was placed in statutory note by Congress); Kostan v. Ariz. Nat’l Guard, 50 M.S.P.R. 182 (1991) (“When the instant case was appealed to the Federal Circuit, the intervenor [National Guard Bureau] brought to the attention of the court an uncodified provision of the National Guard Technicians Act that was not considered in Gordon.”); Appellant’s Opening Brief, United States v. Morales-Hernandez, No. 18-10491 (9th Cir. May 11, 2020), 2020 WL 2543989, at *8 (“The Karingithi Court overlooked—because neither party addressed and the relevant law is not codified (or searchable) in Title 8 of the United States Code—that Congress did in fact address Immigration Court jurisdiction [in a statutory note].”); Brief of Appellee, Nat’l Home Eq. Mortg. Ass’n v. Face, 239 F.3d 633 (4th Cir. 2001) (Nos. 99-2331, 99-2386), 1999 WL 33613774, at *8 (“Defendants rely on an uncodified note appearing at 12 U.S.C. \textsection{3801} to urge this court to strike down the OTS regulation, to refuse deference to its opinion, and to conclude the Virginia law governs AMTs with respect to limits on prepayment penalties.”); Amandeep S. Grewal, Legislative Entrenchment Rules in the Tax Law, 62 ADMIN. L. REV. 1011, 1045 (2010) (explaining that the Tax Court in Cap. One Fin. Corp. v. Comm’r erroneously dismissed taxpayer’s arguments based on an assumption that a provision in the Tax Code nullified provisions in a statutory note).
\item 144. Youmans, supra note 70, at 590.
\item 145. As it was before there was an official code, see, e.g., 1 FEDERAL STATUTES ANNOTATED ii (William M. McKinney & Charles C. Moore eds., 1903) (“The notes, constituting the matter of the smaller type, are of course, the work of the editors.”).
\end{itemize}
Inconsistency

¶45 First, the classification process does not allow for a definitive categorization of the types of provisions that appear as statutory notes. The Office of the Law Revision Counsel’s website indicates that there are certain types of statutory notes normally classified as such, including “effective dates, short titles, savings, and statutory construction. Statutory notes also include provisions that are somewhat less than general or less than permanent, but still relate to existing Code sections, such as those requiring studies and reports, implementation of regulations, or the establishment of a task force.”146 While these types of provisions may normally be classified as statutory notes, the lack of a concrete rule leaves researchers in the precarious position of not knowing when a particular provision will be in a code section or a statutory note. This is not to say that OLRC does not apply its rules consistently. Classification is a complicated process that requires the application of many rules and precedents, and ultimately difficult decision-making. The inconsistency discussed here refers to the simple fact that statutory provisions of the same type can be found in both code sections and statutory notes throughout the Code.

¶46 There are many examples of each type of statutory provision listed above being found in both code sections and statutory notes, and in both positive and nonpositive law titles.147 Title 29, for example, requires a study by the Secretary of Labor in § 624 but also in a note to § 622.148 The Consumer Review Fairness Act (15 U.S.C. § 45b) is a good example of a statutory section that contains rules of construction, savings provisions, and an effective date, in addition to substantive law, in a code section.149 Additionally, 15 U.S.C. § 45b contains definitions, while 15 U.S.C. § 45c’s definitions are in a statutory note. This lack of consistency in location is one of the most vexing research problems surrounding statutory notes. Researchers must not only learn of the existence of statutory notes, but must also deal with the fact that there is no consistent rule regarding which types of provisions end up as statutory notes.150

¶47 The recent reclassification of title 34 of the United States Code illustrates this problem. In an effort to clarify the law in the area of crime control and law enforcement, the OLRC moved provisions from titles 18, 28, and 42151 to an empty

146. About Classification, supra note 61.
150. This article does not head down the difficult road of determining which types of provisions, if any, should end up as statutory notes. It is very likely that some provisions, like short titles, have a place as statutory notes, but a broader discussion is needed to determine the benefits and drawbacks of such an approach. One possible approach is that followed by the Code of Federal Regulations, which only publishes regulations “having general applicability and legal effect.” 44 U.S.C. § 1510; see also 1 C.F.R. § 8.1 (2020).
151. Titles 18 and 28 are positive law titles; title 42 is a nonpositive law title.
In doing so, the OLRC was able to move 41 statutory notes into sections of the United States Code. These former notes comprised a variety of types of provisions, including findings, rules of construction, severability, definitions, and reports. This should be seen as a positive step for researchers, who can now find these provisions more easily. At the same time, however, a number of statutory notes from title 42 remained as such, despite their similarity to other notes moved to sections in title 34.

Researchers also confront inconsistency when acts that are obvious candidates for Code section placement, such as the Torture Victim Protection Act, do not directly amend positive law titles. Researchers looking for the Federal Advisory Committee Act, the Inspector General Act of 1978, or the Ethics in Government Act of 1978 shouldn’t have to consider whether these important laws are in Code sections or an appendix to the Code, but they do. The mere fact that some substantive acts of general application such as these can end up in statutory notes or appendices should worry researchers. Luckily, after allowing these laws to languish for 40 years outside Code sections, the OLRC recently proposed a draft bill that would transfer these three acts into Code sections in title 5. Other important acts have not been so lucky. Unfortunately, as is the case with many positive law bills, this bill remains stalled in the House Committee on the Judiciary.

Another reason statutory notes get lost is that they are lumped together with editorial notes. The lack of a clear distinction between the two makes them appear to be of the same value. While there is a pattern to identify statutory notes, it is not intuitive or clear. This makes statutory notes less findable for all but extremely well-informed researchers. Compounding this is the fact that editorial notes appear before statutory notes. Researchers who read past the statute will find headings such as “Historical and Revision Notes” or “References in Text,” which provides little incentive or warning that statutory text may lie ahead. In some cases, editorial notes are extensive and, at first glance, provide similar information to statutory notes. The editorial notes to 8 U.S.C. § 1101, for example, include “references in text,” “codification,” and “amendment” notes and span eight pages of the print Code before they give way to statutory notes. At least five of these pages are

—

152. Title 34 had been vacant since it was repealed in 1956, and its contents regarding the Navy moved to title 10. See An Act To revise, codify, and enact into law, title 10 of the United States Code, “Armed Forces”, and title 32 of the United States Code, entitled “National Guard,” ch. 1041, 70A Stat. 1 et seq. (1956).
153. 34 U.S.C. § 30501; id. § 40902.
154. Id. § 30506.
155. Id. § 30505.
156. Id. § 40903.
157. Id. §§ 41301, 41306, 40721.
158. See, e.g., id. § 10651 note (findings); id. § 10321 note (statutory construction); id. § 40701 note (reports to Congress).
160. See supra note 113 and accompanying text.
amendment notes that quote portions of public laws to tell researchers what has been added or deleted over the years, making them look quite similar to statutory notes. All of this, in addition to the fact that there is no specific heading to label the break between editorial and statutory notes, makes it more difficult for researchers to appreciate the importance of statutory notes and to discover them.

¶50 Another problem with placement is that the Code sections to which statutory notes are attached do not always provide legal researchers with notice that statutory notes may be lurking. Much of this problem has to do with freestanding provisions that may or may not relate closely to sections of the Code that already exist. The OLRC does its best to place statutory notes with relevant code sections, but it can be difficult to find a fit that is useful to researchers. Because of this, OLRC often adds statutory notes to the first section of a statutory chapter. This causes two problems. First, these beginning sections can be the home for many statutory notes, making them even more difficult to sift through. For example, 8 U.S.C. §1101 has 52 statutory notes, without counting short title and effective date notes. Second, beginning sections are often definition sections that may be unlikely candidates for housing important statutory notes. The Carriage of Goods by Sea Act, for example, a law that appears frequently in appellate court decisions, is a note to a Code section that contains one definition.\footnote{162}{See 46 U.S.C. §30701.} Inexperienced researchers may not think to check for statutory notes in a common section such as this.

Legal Research Providers

¶51 While some scholars have highlighted the existence of statutory notes, none has looked closely at how legal research systems help or hinder researchers in finding applicable statutory notes. Michael Lynch came the closest in 1997 when he noted that one statutory note he examined was found easily in the United States Code index while another was not.\footnote{163}{Lynch, supra note 10, at 79–80. Lynch ultimately concluded that determining whether the index was a good locator of statutory notes would take a larger study. Id. at 80.} Since that time, legal research has moved primarily online, and the way that legal research vendors present information may be less consistent across systems than it was in the days of print. This is especially true as more legal research systems enter the market. Examining how various legal research systems allow for browsing, searching, and index access to statutory notes can help alert researchers to limitations in certain research systems and can help uncover the best ways to research statutory notes.

¶52 Shortly after President Donald Trump fired FBI Director James Comey and began the search for his replacement, Professor Will Baude\footnote{164}{We credit Baude’s article, William Baude, Foreword: The Supreme Court’s Shadow Docket, 9 N.Y.U. J.L. & L.E.T.S.C.B.O.C.T.B.Y. 1 (2015), as the inspiration for this article’s title.} pointed out that the text of 28 U.S.C. §532 states that “[t]he Attorney General may appoint a Director of the Federal Bureau of Investigation.”\footnote{165}{Will Baude, Opinion, Reminder: The United States Code Is Not the Law, Wash. Post: VOLOKH CONSPIRACY (May 15, 2017), https://www.washingtonpost.com/news/volokh-conspiracy/wp/2017/05/15/reminder-the-united-states-code-is-not-the-law/?utm_term=.dd2a8177d732 [https://perma.cc/8HWB-A5NU].} Puzzled because of news reports that the President was the one with authority to appoint the director, Baude turned to the
notes of 28 U.S.C. § 532 to find Public Law 90-351, which gave the President power to appoint an FBI director, “by and with the advice and consent of the Senate.”166 After alerting readers to this example of the fact that the United States Code does not contain all law, Baude points out that “other than a momentary confusion, nothing much ended up turning on this—especially since you can find the reference in the formal notes to the now-obsolete Section 532.”167 But what if finding the formal notes is not so easy? Baude links to Cornell Law School’s Legal Information Institute website, which has a tab for the statutory text and a tab for notes, making the notes relatively easy to find if you know they’re potentially important.168 The printed official version of the United States Code provides the notes directly after the statutory text, although the mixture of editorial notes and statutory notes can cause confusion for the researcher. Most legal research providers follow one of these options for presenting United States Code notes: tabs169 or after-text.170

¶53 Westlaw is the only major legal research provider that breaks the mold here, but not in a good way. Westlaw’s approach to presenting statutory notes makes them much more difficult to find. Take 28 U.S.C. § 532 as an example. The trend of Westlaw’s approach to statutes since it rolled out WestlawNext has been to clear up the statutory text page. Case annotations (called “Notes of Decisions” in Westlaw) and other editorial content are presented via a variety of tabs. There are benefits to being presented with only the statutory text of a United States Code section. Since “we’re all textualists now,”171 it makes sense for researchers to focus on the text first and then move on to annotations and legislative history. But, as we’ve discussed, sometimes the text isn’t in a United States Code section, but in statutory notes, and these are obscured in Westlaw.

¶54 When presented with a United States Code section like 28 U.S.C. § 532 in Westlaw (see figure 2), there is no indication on the face of the page that statutory notes even exist. Instead, researchers are presented with the section text and a number of tabs that lead to other related documents. There are tabs for “Notes of Decisions” (which most legal researchers should recognize as containing case annotations), “History,” “Citing References,” and “Context & Analysis.” None of these identifiers gives researchers a good idea of whether there are statutory notes and where to find them. Researchers may think “Context & Analysis” is the right tab, but this simply leads to references to secondary sources and applicable Topics and Key Numbers. “History” is, in fact, the tab that eventually leads the researcher to statutory notes—which is far from obvious.

166. Id.
167. Id. That said, putting notes in a tab may make them more difficult for certain users to find.
168. See id.
170. Lexis Advance, Fastcase, Casemaker (adding these are “Notes from the Office of Law Revision Counsel”), Casetext, and Bloomberg Law. At the time of this writing, Ross Intelligence did not include any notes in their United States Code database.
Figure 2
28 U.S.C.A. § 532 as presented on Westlaw
Reprinted with permission of Thomson Reuters

Figure 3
28 U.S.C.A. § 532’s History Tab Drop-Down Menu in Westlaw
Reprinted with permission of Thomson Reuters
§55 Hovering over the “History” tab (see figure 3), researchers are given a list of subsections that includes “Historical Overview,” “Graphical Statute,” “Validity,” “Versions,” “Editor’s and Revisor’s Notes,” “Bill Drafts,” and “Legislative History Materials.” Again, there is no clear winner here. Researchers who understand the existence of statutory notes may select “Editor’s and Revisor’s Notes,” but those researchers are likely few.

§56 Alternatively, researchers can click directly on the “History” tab, which gives a description of each of these categories (see figure 4). This may lead researchers further away from finding statutory notes, however, because the description for “Editor’s and Revisor’s Notes” states this content will allow researchers to “review legislative changes affecting a statute section as detailed by West attorney editors or state revisors’ offices.” That may be true for state codes, but is inaccurate with regard to the *United States Code*. The notes for *United States Code* sections are not created by West attorney editors but by the OLRC. Additionally, while the phrase “Editor’s and Revisor’s Notes” may technically be correct for editorial notes, it’s unlikely to help most researchers who are unfamiliar with the process of the creation of the *United States Code* and the OLRC’s role.

§57 Clicking on “Editor’s and Revisor’s Notes” takes researchers to a reproduction of the Historical and Revision Notes that are included with the official *United States Code*, which includes statutory notes (see figure 5). Interestingly, at this point Westlaw does more accurately describe them as “Historical and Statutory Notes,” after the more general “Editor’s and Revisor’s Notes” heading.

---

172. While the notes for 28 U.S.C. § 532 in Westlaw are a replica of the print, many other note sections in Westlaw add legislative reports before the official notes. See, e.g., 8 U.S.C. § 1101 note.

173. After this article was posted on SSRN, Westlaw did away with the label of “Historical and Statutory Notes” and began using more specific labels, such as “Historical Notes,” “Statutory Notes,” and “Savings Provisions,” to refer to different types of notes. We applaud Westlaw for moving in the direction of trying to clarify this confusing research area. We note, however, that there is still a
¶58 Understanding the importance and existence of statutory notes can be difficult enough when this content appears directly after a Code section, but when they are hidden from view, as they are when browsing the United States Code Annotated in Westlaw, researchers who do not know about statutory notes are at a serious disadvantage. Even researchers familiar with statutory notes may find it difficult to locate them or to remember to check for them when they are buried within Westlaw. Missing a statutory note can mean missing a relevant and meaningful provision of law. As detailed below, Westlaw should consider some revisions to its presentation of statutory notes to better assist legal researchers.

Research

¶59 Federal law researchers should be familiar with statutory notes and how to find them. Unfortunately, legal researchers have been given little guidance in locating statutory notes. No current legal research text that the authors are aware of discusses strategies for finding statutory notes.174 Most legal research texts today don’t mention ways to go as some categories, such as “Savings Provisions,” are themselves statutory notes and should be labeled as such.

174. Marjorie Dick Rombauer’s Legal Problem Solving, last published in 1991, provides this guidance: “The variety of sometimes crucial information that may be provided in these fineprint notes is such that you should always scan all the notes to consider possible relevance to your problem.” Marjorie Dick Rombauer, Legal Problem Solving 232 (5th ed. 1991).
statutory notes at all, and those that do touch on the topic do so only briefly. In the balance of this article, we enumerate certain general principles that legal researchers should be familiar with to research statutory notes. We then examine useful research tools, including secondary sources, statute compilations, searching, and indexes that should be used in statutory notes research, and identify limitations.

**General Principles**

¶60 With a research subject as unique as statutory notes, it’s useful to begin with a summary of general principles, which have been explored more fully above. Researchers who understand these principles will be better prepared to identify and find statutory notes in their research.

*Statutory notes exist and are law.*

¶61 Researchers must understand that not all notes in the *United States Code* are editorial. Some contain law that may be relevant to their case. The placement of a provision of law as a statutory note does not affect its validity.

*Statutory notes are identifiable—if you know what to look for.*

¶62 While difficult to spot at first glance, statutory notes do have a somewhat consistent pattern that researchers can use to identify them. They are often introduced by a unique heading that describes their public law of origin. They then identify the citation beginning with the public law number, followed by “provided” or “provided that,” and then a quotation of the statutory provision. For example, *Pub. L. 102-256, Mar. 12, 1992, 106 Stat. 73, provided that: “ . . . .”*

*Statutory notes appear in both positive and nonpositive law titles.*

¶63 While statutory notes tend to appear more often in positive law titles, there are many statutory notes in nonpositive law titles as well.

*All types of provisions can be Code sections or statutory notes.*

¶64 Researchers should not be fooled into thinking certain types of provisions always show up in either statutory notes or Code sections. Because of historical preference or a focus on act-Code coherence, there is no way to say definitively that a certain type of provision always shows up as a statutory note or a Code section.

*Know your research service and helpful research sources.*

¶65 The presentation of statutory notes differs from vendor to vendor. Researchers must be aware of how their provider displays (or hides) statutory notes. Researchers should also be familiar with research sources that help identify statutory notes.

**Secondary Sources**

¶66 Because statutory notes are in the shadows of traditional code research, secondary sources can be an extremely valuable way to find some notes. Secondary sources

---


sources provide researchers with analysis of a particular research topic and citations to relevant primary sources. While the depth and quality of individual secondary sources varies, experts regularly suggest beginning research with these sources to leverage work that has already been done. Secondary source authors are often experts in their fields, which means they should be familiar with the nuances of the topic and legislation that affects it. These qualities lend themselves well to identifying and being familiar with provisions found in statutory notes.

Information on the Torture Victim Protection Act, for example, can be found in multiple secondary sources. Treatises like Wright and Miller’s *Federal Practice and Procedure* and *Litigating International Torts in U.S. Courts*, legal encyclopedias like *American Jurisprudence, 2d* and *Corpus Juris Secundum*, and traditional research tools such as *Causes of Action* and *American Law Reports* all discuss the Torture Victim Protection Act. Researching by subject in these secondary sources would lead researchers to this important cause of action and alert them to the fact that it is found in a statutory note. This would likely be a much quicker solution to finding this statutory note than searching the Code.

While secondary sources can be useful tools for finding statutory notes, they are not without limitations. Some topics of interest simply are not covered in relevant secondary sources. Additionally, certain statutory notes that are only part of a larger act, such as implementations of regulations or rules of construction, may not be singled out in a secondary source. So, while researchers should use secondary sources to be directed to some statutory notes, they cannot rely on them entirely to direct them to all statutory notes. Researchers should be familiar with important secondary sources in their research area and use them when possible.

**Statute Compilations**

Another useful research tool to find some statutory notes is a statute compilation. Statute compilations are unofficial documents that “incorporate[] the amendments made to the underlying statute since it was originally enacted.” In other words, these statute compilations start with a complete statute and bring in all subsequent amendments to make the underlying law current. They are distinct from the *Statutes at Large* in that they meld multiple public laws to state only the law that is currently in force. They are different from the *United States Code* in that the law is kept together rather than scattered throughout the Code and the original title and session law section numbering is retained. Statute compilations can be thought of as an intermediate step between the *Statutes at Large* and the *United States Code* and are primarily created for major acts otherwise found in nonpositive

---


law titles of the Code. The Consumer Product Safety Act, for example, has been amended many times since its enactment in 1972. Looking at the Statutes at Large, a researcher would have to find the original Act and all later amendments in various volumes of the Statutes at Large and then piece them together. A statute compilation brings these changes together. Statute compilations are produced by several different groups, including the House Office of the Legislative Counsel (HOLC), federal agencies, and private publishers.

¶70 One benefit of using a statute compilation in researching statutory notes is that provisions of law that were part of the original act that were moved to statutory notes are returned to their original placement. For example, § 406 of the INA contains a separability provision and § 407 contains an effective date provision that were both classified as statutory notes to 15 U.S.C. § 1101. In the HOLC’s statute compilation of the INA, both of these provisions are included as sections with a reference to where they are found in the United States Code (see figure 6).

Figure 6
Portion of Immigration and Nationality Act—HOLC

<table>
<thead>
<tr>
<th>SEPARABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEC. 406. [8 U.S.C. 1101, note] If any particular provision of this Act, or the application thereof to any person or circumstance, is held invalid, the remainder of the Act and the application of such provision to other persons or circumstances shall not be affected thereby.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EFFECTIVE DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEC. 407. [8 U.S.C. 1101, note] Except as provided in subsection (k) of section 401, this Act shall take effect at 12:01 ante meridian United States Eastern Standard Time on the one hundred eightieth day immediately following the date of its enactment.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 2—REFUGEE ASSISTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OFFICE OF REFUGEE RESETTLEMENT</td>
</tr>
<tr>
<td>SEC. 411. [8 U.S.C. 1521] (a) There is established, within the Department of Health and Human Services, an office to be known as the Office of Refugee Resettlement (hereinafter in this chapter referred to as the “Office”). The head of the Office shall be a Direc-</td>
</tr>
</tbody>
</table>

¶71 Perhaps an even more useful research feature is that statute compilations often refer researchers to statutory note provisions that are not part of the underlying act but are closely related. In essence, these notes serve a similar function to statutory notes in the United States Code, but statute compilations generally make these notes easier to find and more clearly show that they are valid law. The HOLC’s INA compilation, for example, includes the provision that affects the interpretation of 8 U.S.C. § 1101(a)(27)(I) and is included as a note to 8 U.S.C. § 1101, as shown in figure 7. This INA compilation provides a footnote at the beginning of the sec-

180. Of course, unlike the Statutes at Large or the United States Code, these statute compilations are not official.
182. Id. at 37–38.
184. These notes are often part of an amendment to the underlying public law but do not amend the public law directly.
tion alerting the reader that this change came from the Immigration Technical Corrections Act of 1988. Because the footnote is placed with the relevant subsection instead of at the end of an extremely long section, the researcher can more easily understand the applicability of the note. The note also stands alone rather than being encumbered by a number of other editorial and statutory notes.185

¶ 72 In this example, the statutory note applies directly to the INA provision referred to; in other instances, one provision of an amending public law may directly amend the act while other provisions do not. These additional provisions are added as statutory notes to the United States Code, and some statute compilations will include them near the added provision. In the example in figure 8, the researcher is alerted to the fact that subsection (h) was added to the Securities and Exchange Act of 1934 by a certain provision of Public Law 103-202. The compilation footnote then explains that other related provisions—set out as statutory notes in the United States Code—were part of the same public law and provides the text of these provisions and their United States Code citations.

¶ 73 Researchers should be familiar with statute compilations and use them when available. But, like treatises and topic guides, statute compilations have some limita-

Figure 7
Immigration and Nationality Act—HOLC

Figure 8
Securities and Exchange Act of 1934—HOLC

185. See supra notes 161–62 and accompanying text.
tions. First, statute compilations are not created for acts found in positive law titles of the Code. Therefore, research on statutory notes in positive law titles—a majority of notes found in our survey of notes cited in federal appellate courts—would not be benefited by looking at statute compilations. Second, statute compilations are most often created for large, complex laws that are amended frequently. While this is extremely helpful for acts such as the INA or the Social Security Act, smaller acts such as the BOTS Act are unlikely to have their own statute compilation. Third, researchers must know where they need to look before statute compilations can help them find statutory notes—they must know in which haystack to look before effectively searching for their needle. This makes statute compilations quite valuable for researchers who know they will be looking amidst the INA or the Securities and Exchange Act, but for researchers who aren’t that far along, statute compilations may be of less help initially. Finally, not all statute compilations are created equal. While the text of the acts are the same across statute compilations, the statutory notes that are included can be different. For example, as shown in figure 7, the HOLC version of the INA contains a statutory note pertaining to § 101(a)(27)(I). The American Immigration Lawyers Association’s INA compilation, popular among practitioners, does not include this statutory note. It does, however, include a statutory note relevant to § 101(a)(15)(T), shown in figure 9, that the HOLC version does not. The differences between statute compilations mean researchers cannot be assured that they are seeing all relevant statutory notes when using a statute compilation.

Searching

Searching the United States Code in the various commercial research systems does yield results from statutory notes in addition to section text. A researcher looking for information on the appointment of the director of the FBI could run a
search for something like director /s “federal bureau of investigation” /s appoint! and be led to 28 U.S.C. § 532 within the first few results, depending on the system. Because the statutory text is followed by the notes, a researcher can look for the highlighted search terms and see them in full. This is true even on Westlaw, which does not initially show the notes with the statute. The fact that running a search does bring back statutory notes in search results, however, does not alleviate the problem that arises if a researcher does not understand the importance of statutory notes. Here, for example, the highlighted search terms show up in the text of 28 U.S.C. § 532 as well as in the notes. Unless a researcher understands the importance of statutory notes, she may never venture below the code section text.

¶75 However, statutes don’t always lend themselves to effective searching. Statutory language is often technical, and a legal researcher often must have the precise language to find relevant results, especially if statutory searching is not aided by annotations. Statutes are usually situated in a statutory scheme that is most easily explored by browsing. Because statutory notes are not always directly related to the code section they are associated with, it isn’t enough to be led by a search to one code section and then browse the rest of the statutory scheme to discover the law. More often, researchers must hit a statutory note precisely with a search or they will miss it altogether without other research strategies.

Index

¶76 While searching has become the preferred method of legal research these days, there is something to be said for the use of indexes when looking for statutory notes. Because of the technical language used in statutes, legal research textbooks consistently encourage the use of indexes when dealing with statutory research.

188. Of course, among the many providers, the exact Boolean symbols for truncation, proximity, etc., will slightly vary.
189. When research for this article began, a search in Westlaw retrieved only portions of the statutory notes that contained the search terms. The note provisions were divorced from any context and gave little guidance about where they came from. Additionally, paragraphs with matching search terms were simply listed one after another with no separation between the ending and beginning of different public laws. For example, the search referenced in this paragraph provided hits in six different paragraphs that were listed one right after another. We applaud Westlaw for improving its presentation of search results that include statutory notes.
190. See id. at 15 (“The most efficient way to locate statutes pertaining to a particular set of circumstances is through a code index. Thus, the very first step in most legal research using primary materials is to consult the index of the appropriate statutory code.”); Mersky & Dunn, supra note 28, at 177; Olson, supra note 175, at 63 (explaining that despite some flaws, “an index may nonetheless allow you to zero in on statutes that are directly on point more quickly than a full-text keyword search”); Amy Sloan, Basic Legal Research 172 (6th ed. 2015).
Statutory notes are no exception to this. As mentioned above, the General Index to the *United States Code* contains more than 44,000 entries that reference statutory notes.\(^{194}\) These entries can often be much more accessible to a researcher than trying to construct a search. For example, instead of coming up with a precise search, a researcher could go to the index and find a section for “Federal Bureau of Investigation” and a subsection for “Director,” under which he or she will find a number of citations to relevant *United States Code* sections. Entries for “compensation,” “confirmation,” and “succession to office,” among others, provide citations to statutory notes.

\(^{77}\) Unfortunately, most legal researchers will not have access to the General Index of the *United States Code*. Unlike the Code itself, the General Index is created and supplied by Thomson Reuters and is not available for free online.\(^{195}\) Westlaw does provide a version of this index. Using it, however, requires a knowledge of how to find statutory notes on Westlaw as described previously.\(^{196}\) This is because when researchers click on an index entry such as “Federal Bureau of Investigation – Director – Confirmation: 28 USCA § 532 NT,” they are taken to 28 U.S.C.A. § 532, not to its notes. Researchers with little knowledge of statutory notes may not even think to look for statutory notes. Lexis also provides online indexing to its *United States Code Service*,\(^{197}\) but its index references suffer from the same problem of linking researchers to the text of the section rather than to the statutory note.

### Improving Research Access

\(^{78}\) Legal research providers need to be aware of the importance of statutory notes and make them accessible to researchers. Westlaw in particular needs to re-envision its presentation of statutory notes so as not to obscure further an already difficult-to-find source. Despite moving notes to a tab, Westlaw still keeps presidential memoranda visible beneath statutory sections.\(^{198}\) At the very least, statutory notes should appear here, too. Westlaw could also consider adding a new tab for editorial and statutory notes, or at the very least renaming the current “Editor’s and Revisor’s Notes” to “Editorial or Statutory Notes.” This could also include an entirely new way of imagining how statutory notes are portrayed.

\(^{79}\) The OLRC could also help researchers by more clearly labeling editorial and statutory notes. Some division already exists, with editorial notes generally appearing directly after a statutory section and statutory notes following editorial notes. Neither is labeled as such, however, making it difficult for researchers and legal research providers to quickly see a distinction. Statutory notes, for example, are each labeled with a heading that reflects the subject of the public law that is the subject of the statutory note. Clearly labeling the division between editorial notes and statutory notes could help researchers be attuned to the different types of notes.

---

194. See supra note 137 and accompanying text.

195. See 38 *United States Code*, at ii (first volume of General Index, stating that “[t]he General Index material contained in this volume was created and supplied by Thomson Reuters which claims a copyright therein”). HeinOnline, another commercial subscription legal research provider, also maintains online, searchable copies of all official *United States Code* editions, including General Index volumes.

196. See supra ¶53.

197. Lexis’s *United States Code Service* index is very difficult to find and would benefit from being more prominently displayed.

that may affect them. Inclusion of an appropriate tag in the xhtml version of the Code could allow legal research providers to programmatically highlight or extract statutory notes and make them more visible to researchers.

¶80 Will Tress has proposed the creation of an unofficial government-provided electronic version of the United States Code that could be used to experiment with bringing notes into the text of a statutory section. One of the current limitations to such an idea is the fact that it is extremely difficult even to identify statutory notes in a programmatic fashion. Clearly labeling statutory notes as such would allow legal research providers to identify these important notes at a much lower cost than currently. Reducing that cost could incentivize legal research providers to begin experimenting with ways to display statutory notes in a more useful manner. While we support Tress’s ideas of experimentation in an unofficial United States Code, we find it more plausible that commercial legal research providers would be in a better position to pull this off.

¶81 While these proposals would help alleviate some of the frustration in finding statutory notes, they ultimately do not fix the Code’s underlying structural problems that lead to statutory notes. More sustained discussion is needed to address these problems. At first glance, it may seem like the best solution is to push for increased positive law codification. However, despite OLRC’s efforts, this appears to be a low priority for Congress. While an uptick in positive law codification would help the Code move away from the bizarre split system between positive and nonpositive law it currently occupies, it would likely lead to even more statutory notes. The upcoming centennial of the United States Code’s creation in 2026 may provide an opportunity to evaluate where the Code should be moving in its next 100 years.

Best Practices for Teaching Law Students About Statutory Notes

¶82 Researchers and lawyers should follow the guidelines listed above, but legal research teachers, law librarians, and others providing practical instruction on legislation should help their students develop comfort in this area. As law librarians instruct students on statutes and statutory research, instruction about statutory notes should become part of the canon. Teachers must acknowledge the complexities of statutory notes research, making space for it in the legal research curriculum.

¶83 Research instruction about statutory notes should address three complex features of the Code. First, students should understand the story of the contemporary United States Code and how it came to be. Next, students should understand the distinction between positive law and nonpositive law and its implications for research. Finally, students need to know what forms of statutory law evade codification in traditional Code sections and how to find those provisions. As they teach, instructors can incorporate each of these concepts into the foundational concepts of statutory research.

¶84 The United States Code in its current form is one of several publications memorializing legislation; it is not the only one, but it is distinct in its topical arrangement. Although the story of codification has a rich history, students need to understand the basic history as described earlier in this article. The story of the
Revised Statutes sets the stage for understanding the distinction between positive and nonpositive law. To underscore this distinction, teachers can present an enacted positive law codification statute. Using this unique form of legislation makes clear the nature of positive law titles: they are the law because the title and all of its underlying sections and subsections have been approved by Congress and signed by the President. Title 54—National Parks Service and Related Programs—makes a clear example of a title recently enacted into positive law. Enacted into positive law by Public Law 113-287, the statute describes the title as follows:

[A] restatement of existing law relating to the National Park Service and related programs as a new positive law title of the United States Code. As with all positive law codification measures, the enactment of title 54, United States Code, did not create new law or change the meaning or effect of existing law. Instead, the organizational structure of the existing law was improved, and ambiguities, contradictions, and other imperfections in the law were removed. Detailed information about Public Law 113-287 is available in the accompanying House Report 113-44.²⁰⁰

Students can locate the bill, committee reports, and floor activity, and review the legislative history of Public Law 113-287 as with any other statute.

¶ 85 Those titles that have not been subjected to this legislative process remain nonpositive law wherein the Code contains only prima facie evidence of the law. For these titles the public laws contained in the Statutes at Large are legal evidence of the law. In practical terms, teachers should underscore that for nonpositive law titles, when there is an inconsistency between the statutory language in the Code and the Statutes at Large, the text in the Statutes at Large governs. Legal research teachers can make this concept clear by using two sources: a federal statute and case law. 1 U.S.C. § 112 provides that “The United States Statutes at Large shall be legal evidence of laws.”²⁰¹ Similarly, Section 204(a) of the same title provides that the United States Code “establish[es] prima facie the laws of the United States, general and permanent in their nature, in force . . . . Provided, however, that whenever titles of such Code shall have been enacted into positive law the text thereof shall be legal evidence of the laws . . . .”²⁰²

¶ 86 There are dozens of cases—and even a Topic and Key Number in the West Digest—that announce this principle. The leading case is United States National Bank of Oregon v. Independent Insurance Agents of America, Inc. In a matter turning on the use and precise placement of quotation marks in an amending statute, the Court made a close interpretation of statutory history to determine whether 12 U.S.C. § 92 (appearing in the 1952 edition of the Code, but noted as repealed in subsequent editions) had been repealed or not, holding:

Though the appearance of a provision in the current edition of the United States Code is “prima facie” evidence that the provision has the force of law, 1 U.S.C. § 204(a), it is the Statutes at Large that provides the “legal evidence of laws,” § 112, and despite its omission from the Code section 92 remains on the books if the Statutes at Large so dictates.²⁰³

²⁰². 1 U.S.C. § 204(a).
²⁰³. 508 U.S. 439, 448 (1993). See also United States v. Carroll, 105 F.3d 740, 744 (1st Cir. 1997). This case involved the placement of a comma in a statute, and there was a discrepancy about
87. After teaching about positive and nonpositive law titles, instructors can help students understand how amending these different types of titles can affect the creation of statutory notes. For example, it is instructive for students to be able to explain why the Torture Victim Protection Act of 1991 is codified in a statutory note: as explained above, title 28 is a positive law title, and reading the enacted bill, students will see that Congress failed to specify where in the Code the Act should be inserted. Students should also be taught the types of statutory provisions that are commonly found as statutory notes in both positive and nonpositive law titles. Instructors must make sure, however, to teach students that there are many exceptions to this. They could refer to the many examples provided in this article (or others that they find) to demonstrate this point.

88. In the classroom, students find the process of codification as well as the research implications of positive law abstract. Asking students to become codifiers and attempt the process using a simple public law can make the process more concrete. Moreover, this type of short exercise brings all the concepts described above together. Using Public Law 116-12, an act to clarify the grade and pay of podiatrists employed by the Department of Veterans Affairs, students codify provisions, create statutory notes, and see the implications of the positive/nonpositive law distinction. Additionally, this exercise shows how one relatively short public law amends several different code sections.

89. To conduct this exercise, first ask students to determine whether the title is positive or nonpositive law. Students can use the OLRC’s website to confirm that title 38 has been enacted into positive law, but they should also note that the Act begins with the words “to amend title 38, United States Code.” These words signal a positive law title. (A law in a nonpositive law title would be amended with reference to the section number from the original public law.) Next, students can move through the Act, section by section, to determine the location of those sections in the Code.

90. Finally, researching statutory notes can be explained by showing how to find any of the examples presented in this article with the research principles explained above. Then students should be asked to solve a research problem that requires them to find a statutory note. Students won’t fully grasp how to conduct the research or the importance (and difficulty) of it until they’ve had a chance to try it themselves.

91. To be sure, requiring students to determine at the outset of their statutory research whether a statute they are searching for may be codified in a nonpositive or positive law title may be pedagogically unrealistic. Instead, students should always consult secondary sources when they are conducting statutory research. Congressional Research Service (CRS) reports can be the first stop for students.

the existence of the comma between the Statutes at Large and the Code. The Court stated, “[such conflicts are] rare, but, when they occur, the rendition of the law contained in the Statutes at Large controls.”


205. This example was provided by OLRC. E-mail from Robert Sukol, Deputy Law Revision Counsel, OLRC, to authors (July 1, 2019, 14:17 EDT) (on file with authors).

206. Id. OLRC states that this statute demonstrates how many Code sections are amended in one Public Law section.
looking to learn more about a statute, its legislative history, and its construction. For nearly all super statutes and many less complex statutes, there is a CRS report. Today, CRS reports are freely searchable and available on the website everycrsreport.com. Additionally, ProQuest, Lexis Advance, Westlaw, and Bloomberg, among other sources, have CRS reports. These are an invaluable source for understanding the nuances of specific statutes.

For statutory research instruction, teachers can emphasize that although full-text searching may be the most sophisticated, it is among the most difficult methods for locating relevant code sections. In other words, full-text searching on its own should be avoided. Using an index for statutory research helps students see the broad range of statutes that could be relevant. Likewise, indexes provide distinct entries for statutory provisions that appear as notes along with traditional code sections.

Conclusion

Statutory notes exist throughout the United States Code for a variety of reasons. As the Code grows in size and volume, the problem of statutory notes will persist. Likewise, the slow pace—and near standstill—of the enactment of positive law titles of the Code by Congress contributes to the complexities and codification challenges facing the Office of Law Revision Counsel as they fulfill their duties. Legal researchers must be familiar with statutory notes—why they exist and how to find them—in order to accurately research federal law. While there is currently no perfect research tool for researching statutory notes, several different tools are available to help researchers find statutory notes. As legal researchers are more aware of statutory notes and take them into consideration in their research process, they can help to bring statutory notes out of the shadows.
A Different Democratic Divide: How the Current U.S. Online Court Record System Exacerbates Inequality*

Lisa von Wiegen** and Shannon M. Oltmann***

Born-digital documents have overall increased access to information, but reliance on online access can be problematic, particularly for court records. The current system for accessing U.S. court records results in a democratic divide, in which citizens cannot access court records that are essential for oversight of this branch of government.

Introduction ......................................................... 257
Court Documents .................................................... 258
The Critical Role of Court Records in a Democracy ............... 259
Current Access to Federal Court Documents ..................... 262
The Complexity of Accessing Court Documents .................. 264
The Digital and Democratic Divides ............................... 267
Possible Solutions to This Democratic Divide ..................... 270
Future Research Directions ........................................ 272
Conclusion .......................................................... 273

Introduction

¶1 The multifaceted nature of digital divides and their implications are well established in research. For example, we know that a reliance on online access to government documents can be problematic for several reasons. Rebecca Kunkel notes that born-digital documents, without proper preservation and maintenance, are unlikely to be accessible long term; she notes that “whether web publishing infrastructure sufficiently ensures that government information remains available over the long term is an open question.”1

¶2 However, one implication has received scant attention: because so many documents are born digital and/or are stored solely digitally, access to court records may have decreased rather than increased in the information age. This, in turn, affects our ability to function effectively as a democracy because open access to court documents is essential to holding the judicial branch accountable.2

* © 2020 Lisa von Wiegen and Shannon M. Oltmann
** Of Counsel, The Winters Law Group, Lexington, Kentucky.
*** Associate Professor, University of Kentucky, Lexington, Kentucky.
Though court records in the United States are public records (with few exceptions), putting court records online does not always increase access to them.

¶ 3 In this article, we introduce and demarcate the problem of an overreliance on online access to court records in the United States. Next, we provide the conceptual lens of a “democratic divide,” which offers a meaningful way to explore this problem in an information society. Third, we offer several ways to redress the democratic divide and restore more accountability and usability to court records. We conclude by suggesting avenues for future empirical research.

Court Documents

¶ 4 The concept of “court documents” and the court record contained therein must be appropriately defined. While much of society has embraced the cultural revolutions wrought by advancing technology—responding by creating language for differing information formats—the U.S. courts have struggled to maintain procedures, rules, and definitions that are relevant and consistent as times and technology change.

¶ 5 The term “document” here indicates (1) the “record” of the court, meaning materials created by the court in the furtherance of litigation and the running of the courthouse, such as dockets, opinions, transcripts, and audio and data recordings; and (2) items that would be produced by plaintiffs, defendants, and prosecutors. While the term “document” may seem insufficient, more encompassing language quickly becomes somewhat confusing and overly cumbersome.

¶ 6 The judicial system has had difficulty finding an all-inclusive term for such items. Federal Rule of Civil Procedure 34 is entitled: “Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes.” It is important to note that many items included in court records are born-digital documents, photographs, electronic data, digital records, and other items that, at first glance, may not seem like “documents.” Rule 34 includes

[A]ny designated documents or electronically stored information—including writings, drawings, graphs, charts, photographs, sound recordings, images, and other data or data compilations—stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form; or (B) any designated tangible things.

The terms “court record” and “court document,” therefore, refer to any item maintained in the court.

¶ 7 It is also important to note that there is a certain ebb and flow to the court record depending upon the stage of the litigation. Depositions are an example of the changing definition of a court document. Often state and federal procedural rules require that original sworn depositions are held by the deposing attorney until they are filed with the court in preparation for trial or other matter where the

4. Id. R. 34 (a)(1)(A)–(B).
court needs to view the deposition. Later, some or all of the depositions may be filed with the court, either with motions or in their entirety in preparation for trial; thus, they will ultimately be filed and become part of the record.

**The Critical Role of Court Records in a Democracy**

§8 Early American law opened the court and public offices to inspection by U.S. citizens. *State ex rel. Colscott v. King* asserted:

They [citizens] must still be held to have a right to ascertain for themselves this fact and to be informed, if they desire, in respect to the official acts of their officers or servants in the management of the business of the county, and for this purpose, at least, they ought to be accorded the right to obtain all the knowledge and information which can be afforded them through an inspection of the records and documents belonging to the public offices of the county.⁶

Furthermore, the court in *Colscott* concluded:

I do not think that any common law ever obtained in this free government that would deny to the people thereof the right of free access to, and public inspection of, public records. They have an interest always in such records, and I know of no law, written or unwritten, that provides that, before an inspection or examination of a public record is made, the citizen who wishes to make it must show some special interest in such record.⁷

§9 While the *Colscott* decision is about access to public records in general (not specifically court documents), it set precedent for public access. Subsequent courts relied on this interpretation, stating that court records are public records and thus should be available to the public. For example, in 1977, the Supreme Court noted: “It is clear that the courts of this country recognize a general right to inspect and copy public records and documents, including judicial records and documents.”⁸

§10 Public access to the courthouse does not mean access only during proceedings. The public is guaranteed a seat within the court and access to documents and materials provided throughout a trial. Indeed, the Supreme Court asserted in 1947 that “what transpires in the courtroom is public property. . . . There is no special prerequisite of the judiciary to enable it.”⁹ Thus, evidence, answers to interrogatories, sworn statements, and motions are also part of the public record.

§11 However, public access is not absolute despite the case law. Later cases acknowledge the openness of the court while noting some limitations. The Court in *Nixon v. Warner Communications* stated that there is only a presumption of access.¹⁰ If a defendant, a litigant, or the court challenges the public’s right to access, then access must be evaluated against certain factors. In criminal proceedings, judicial decisions to prohibit access to criminal records face a greater amount of scrutiny. Such a prohibition must be shown to be based on a compelling interest, and the prohibition must likewise be narrowly tailored to protect that interest.¹¹

---

6. 57 N.E. 535, 538 (Ind. 1900).
7. Id.
10. See Nixon, supra note 8.
Aside from intrinsic interest in court records, there are many significant reasons to value access to these documents. Access to court documents is essential to the democratic process. The judiciary represents one branch of the U.S. system of democracy, and oversight of that branch is primarily possible through the transparency of judicial records. Democracy demands an open and accessible court system. This openness extends not only to those participating within the legal system but also to citizen-stakeholders. The ability of any citizen to view the actions of the judiciary is a cornerstone of American legal history because it is directly related to citizen involvement in the implementation of the laws. The concept is simple: secrecy breeds injustice. Open access to court documents benefits democracy in several ways; David Ardia categorizes these as first-, second-, and third-order benefits.

First-order benefits “impact the functioning of specific court proceedings.” This includes things such as ensuring witnesses are honest, judges are not corrupt, and the government does not engage in unjust prosecutions. As Amanda Conley et al. explain, “citizens are presumed to have a right to inspect [court records] to ensure that courts are exercising their powers not only competently and fairly but also within the limits of their mandate.” In general, first-order benefits improve court functioning and operations.

Second-order benefits “affect the judicial system as a whole”, these include enhancing public understanding, facilitating transparency, increasing accountability, and improving the predictability of the courts. Grayson Barber notes that one of the primary motivations of open government records, generally, is to enhance public understanding of government systems and processes. With greater understanding comes increased transparency and accountability. John Gathegi notes that “the public scrutiny afforded by publication [of court records]

12. We would suggest that noncitizens likewise benefit from access to court records, though scholarship and case law have focused on the benefits to citizens.
14. See Ardia, supra note 2.
15. Id. at 895.
18. Ardia, supra note 2, at 895.
20. See generally Bepko, supra note 16; Conley et al., supra note 16; Anna L. Endter, Authentication of Online State Primary Legal Resources as a Social Justice Issue: The Uniform Electronic Legal Material Act and How It Can Benefit Pro Se Litigants, 31 LEGAL REFERENCE SERVS. Q. 293 (2012); Gathegi, supra note 16; John N. Gathegi, Officially Mandated Disappearing Information: The Legal Depublication Phenomenon, 22 Gov’t Info. Q. 423 (2005); Schultze, supra note 13; Solove, supra note 16.
reinforces in judges a sense of the public accountability of their work and thus contributes to transparency in the judicial decision-making process."²¹ Openness creates greater justice when citizens know how and why courts are ruling. Open records allow citizens to engage, evaluate, and scrutinize the judiciary, thus eliminating a judge’s ability to obfuscate, favor, or engage in arbitrary decisions. Courts that operate in an environment of transparency operate with an awareness that their decisions can be scrutinized and evaluated, thus adding reliability to the decisions and behavior of the judiciary.

¶ 15 When citizens can inspect, read, and copy court records, the judicial system can be held accountable to the people it serves. In Press-Enterprise Co. v. Superior Court, the Court noted that “[o]penness thus enhances both the basic fairness of the criminal trial and the appearance of fairness so essential to public confidence in the system.”²² Finally, open court records mean that the system is more predictable.²³ This allows people (both lawyers and average citizens) to anticipate how courts might rule. In describing the relative unavailability of court records in Kenya, Gathegi notes that better access enables the doctrine of stare decisis.²⁴

¶ 16 Third-order benefits “influence society broadly”²⁵ and “play an important structural role in our constitutional system by increasing citizens’ ability to exercise self-governance.”²⁶ These benefits include “the application of the law and the mechanisms by which it is formed.”²⁷ Understanding the judicial process is essential because it is one of the foundational branches of government in democracy. Anna Endter explains that “our democratic society rests on the presumption that citizens have a right to access the law that governs them if they are to be meaningful participants in the democratic process”;²⁸ an informed citizenry is one of the key components of an effective democracy.²⁹ As Ardia notes, “public access to the courts helps to mitigate the informational asymmetry that exists between citizens and the government.”³⁰ While citizens have direct recourse (through voting) to communicate with the legislative and executive branches, they have less direct recourse with the judicial branch (generally speaking); open access to court records arms citizens with knowledge about how the laws are interpreted and applied.

¶ 17 The idea that democratic ideals and traditional ideas of a fair judiciary are supported by open access to court documents can be seen by examining the modern-day transformation of the court system in Kenya. Maya Gainer explains that the Kenyan court system was a closed entity to citizens until 2011 and the appointment of Chief Justice Willy Mutunga. Cases languished for years with little citizen oversight. Mutunga, himself a civil activist, was once held for 16 months without trial.

---

22. 464 U.S. 501, 508 (1984); see also Bepko, supra note 16, at 971.
23. See, e.g., Ardia, supra note 2; Gathegi, supra note 16.
25. Ardia supra note 2, at 895.
26. Id. at 897.
27. Id. at 898.
28. Endter, supra note 20, at 296; see also Barber, supra note 19; Solove, supra note 16.
29. Whether citizens choose to become informed is, of course, another matter. Ardia argues, however, “although we might be skeptical that the public will make proper use of the information it gains from observing the courts, this does not negate the need for access in the first place.” Ardia, supra note 2, at 902.
30. Id. at 900.
under a former president. Court documents routinely disappeared. Court cases were adjourned or dismissed at whim instead of through procedural or legal means. A common saying among Kenyans was “Why hire an attorney when you can buy a judge?” The judicial culture was one of “unaccountability, distance, hierarchy, and opacity.”

¶18 To create a new culture and change how citizens saw the judiciary, the new chief justice insisted on creating public access to the court and court documents. This was considered the first of four pillars in the process of reform. While the changes are newly implemented and still evolving, it does appear that such changes, particularly related to public access, have benefited the democracy of Kenya.

¶19 First-, second-, and third-order benefits derive from open, publicly accessible court records. These contribute to the smooth, effective functioning of a democracy, enabling informed and engaged citizens. Access is about the public’s ability to oversee the process of their own judiciary and ensure that the courts are acting within the accepted framework of democracy.

Current Access to Federal Court Documents

¶20 Electronic government, or e-government, is “the use of technology, particularly the Internet, to enhance the access to and delivery of government information and services to citizens, businesses, government employees, and other agencies.” E-government is mentioned here because the provision of court documents online can be seen as yet another manifestation of this movement, though court records have received scant attention in the e-government literature (which focuses mainly on the executive or legislative branches).

¶21 Historically, public access to court records was accomplished by going to a location in a specific jurisdiction, either the courthouse or a clerk’s office, and locating the appropriate file. Often these records would be in print, filed in filing cabinets or on moveable shelving, maintained by office staff, and available upon request. Those who wanted to examine court records could do so for free, though they may have incurred charges for copying or faxing documents. Though access was essentially free, courthouses did have other requirements for those seeking court documents. In some instances, individuals would be required to provide their identification upon entry and pass through a security screening. Shelves of records were also generally housed in a room where clerks assisted the public in

32. Id. at 4.
33. Id.
34. See generally Ardia, supra note 2.
36. See Schultze, supra note 13, for a further discussion of historical context.
38. See, e.g., Conley et al., supra note 16.
locating a specific file. Thus, one could not simply peruse the shelves of courthouse filings.

¶22 In contrast, access to federal court documents today relies upon the Public Access to Court Electronic Records (PACER) system. It is utilized only in the federal court system. PACER is an electronic public access service that allows users to obtain case and docket information online from federal appellate, district, and bankruptcy courts, and the PACER Case Locator. PACER is provided by the Federal Judiciary in keeping with its commitment to providing public access to court information via a centralized service.39

¶23 When PACER was developed in the late 1980s, Congress determined it should be funded by user fees.40 With a credit card, users may access information immediately; those who do not have a credit card or do not want to submit that information must wait to receive an activating code in the mail seven to ten days after requesting it. Currently, these costs are ten cents per page to access a file through PACER, with a maximum charge of $3.00 per document. Fees are waived if the user does not exceed $30 in fees per quarter.41 Court filings can easily reach hundreds of pages per case, and someone researching prior relevant cases might have to access many such records.

¶24 Within federal courts today, everything is essentially electronic. The exceptions for electronic filing are often dealt with by the local rules in each of the separate federal circuit or district courts. Such is the case, for example, in the United States District Court for the Central District of California, which indicates certain circumstances when electronic filing is not necessary.42

¶25 In 2007, the Administrative Office of the U.S. Courts began a PACER pilot program designed to facilitate access to court documents. The program allowed free access to PACER documents at 17 libraries around the United States. Susan Lyons describes how patrons quickly benefited from this access. As with many successful e-government initiatives, the library was a positive force in promoting access and assisting users.43 Lyons indicates that a “steady stream of visitors” utilized the service, including pro se litigants, attorneys, students, and out-of-town journalists.44 In total, an average of 30 patrons a month utilized the service at Lyons’s library.45 After roughly a year of the free PACER program, an open records advocate utilized one of the pilot library’s passwords and log-in information to download 19,856,160 pages (roughly 20% of the total PACER archive) over the span

42. These exemptions include pro se litigants and unusual exhibits, such as items that are not paper-based. See, e.g., Local Rule 5-4.2 (C.D. Cal.), https://www.cacd.uscourts.gov/sites/default/files/documents/LocalRules_Chap1.pdf [https://perma.cc/Y4CV-DHA7].
44. Id. at 31.
45. Id.
of one month. Upon noticing the amount of usage, the Administrative Office of the Courts discontinued the program. The Administrative Office of the Courts provided the information to the FBI, who investigated but declined to prosecute. No program offering free access to PACER has been offered since.

Recent efforts to bring free public access to federal court records are very limited. A pilot project involves telephonic access to the bankruptcy court records. The interactive system provides free public access to bankruptcy case information in English and Spanish, including debtors’ names, methods of case disposition, bankruptcy chapter, reported assets, case status, and discharge date.

The Complexity of Accessing Court Documents

Similar to the digital divide, access is more complex than it may appear at first. It can be defined as “the presence of a robust system through which information is made available”; here, “system” should be understood as the socially and politically contextualized means by which individuals obtain information. In addition, there are physical, intellectual, and social aspects to information access.

Physical access includes “the physical structures that contain information, the electronic structures that contain information, and the paths that are traveled to get to information.” Geography, technology, and economics can all affect physical access. Intellectual access refers to understanding the information in a document, which brings into play traits such as physical or cognitive (dis)abilities, language competence, and technological literacy; whereas physical access is enhanced, constrained, or manipulated in the external environment, intellectual access is affected by an individual’s internal characteristics. Finally, the concept of social access suggests that elements of one’s social world, including social norms and worldviews, influence which information one accesses, and how and why particular information is sought. Thus, neither physical nor intellectual access can be understood in isolation; both are mediated by the social milieu of individuals. Access is most often conceptualized in terms of physical access, but intellectual and social components should be investigated as well, to fully analyze how access may be affected. All of these aspects of information access are implicated by the current system of access to court documents. Below, we discuss physical and intellectual access together because they are often entwined.

---

46. Id. at 31–32.
47. Id. at 31.
52. Id. at 57.
53. See generally id.; Jaeger & Thompson, supra note 35.
In terms of physical and intellectual access, PACER often functions as a digital wall that holds citizens at arm’s length from the documents they are legally entitled to access. In the predigital era, people could access and review court documents without technological devices (though we note that the requirement of physical presence at a courthouse or clerk’s office may have impeded some citizen access). In contrast, PACER users must have technology—some sort of computing device and Internet access. They must be able to locate the service online and then register to use the service. The log-in is the first barrier: there are many members of groups that are leery of utilizing e-government. They are even more concerned when they must provide information about themselves.

As mentioned above, there are user fees to access PACER records; the requirement of a credit card (or a delay while waiting for a mailed access code) can create tiered levels of users. Those lacking a credit card are not able to get information in a timely manner. This may be especially problematic for those from lower socioeconomic backgrounds. As Darrell West explains, user fees can create a two-tier society “based on those who can afford information and those who cannot.”

The second level of physical and intellectual impediments to utilizing PACER is the cumbersome and confusing interface. It is not designed for the unskilled user. Assumptions about users’ skill levels are often problematic and can

---

lead to a failure of physical access. After registration, the log-in is fairly simple, as shown in figure 1.

A user must then select the court of interest (such as the First Circuit Court of Appeals or a U.S. bankruptcy court), as shown in figure 2. Nothing guides the user here; she must already know which court she is looking for.

Figure 2
Structure of Federal Courts as Shown in PACER System

![Structure of Federal Courts as Shown in PACER System](image)

After clicking through another screen, the user may then use the navigation bar across the top to select query, specific reports, or utilities (as shown in figure 3).

Figure 3
Sample Page of PACER System

![Sample Page of PACER System](image)
These are likely not terms that the average user understands in relation to court documents.

¶ 32 In addition to these problems with physical and intellectual access, there are difficulties with the social aspects of access to court documents. First, it seems commonsense that court documents would be physically available in courthouses and clerks’ offices, yet this may no longer be true. Indeed, access at the courthouse is often through a computer terminal located within the clerk’s office. The relocation of court documents to the Internet may be disconcerting to some, especially those whose worldview does not take into account the prevalence of the Internet. Those of lower socioeconomic backgrounds, who may lack familiarity with the Internet, may be particularly thrown off by this change.

¶ 33 A second aspect of social information access concerns use of PACER in public libraries. In most U.S. public libraries, (nearly) every resource is available free of charge. Through local, state, and tax funds, public libraries provide a wide array of resources at no further cost to their patrons, especially government documents. In this respect, PACER is a glaring exception.

The Digital and Democratic Divides

¶ 34 The existence of a digital divide has been well established, though the definition of this concept has shifted over time. It originally referred to differentiated access, the “haves” and the “have-nots.” John Carlos Bertot notes that the digital divide incorporates dimensions of technology, telecommunications/broadband availability, economics, information access, and information literacy.56 Jan van Dijk describes four aspects of access implicated in digital divide studies: motivation access (whether someone wants an Internet connection), material access (having a computer and Internet access), skills access (having the skills needed to use the Internet), and usage access (how and for what purposes the Internet is used).57 Dmitry Epstein et al., in their study of policy frames around the digital divide, collapse these differences into “access” and “skills.”58 JungAe Yang and Maria Grabe describe “gaps in technology usage skills, digital media use, and information acquisition” as a first-level digital divide and a “usage gap” as a second-level divide.59 They explicitly draw attention to the “lack of consensus about [a] conceptual definition” of the digital divide(s).60

¶ 35 Furthermore, Fadi Hirzalla et al. explain that digital divide studies sometimes examine differences between countries and sometimes differences within (usually developed, democratic) nations; they also note that the latter studies generally differentiate between “two aspects of internet use: intensity of internet use and forms of internet use.”61 Finally, Pavle Sicherl and Vesna Dolnicar et al. argue that

60. Id.
how a divide is measured matters: the type of measurement (and statistical analysis) chosen affects the data that is produced and the interpretations that can be supported by the data (including policy framings).62

¶36 In summary, “digital divide” seems to be used in multiple, overlapping, sometimes competing or conflating ways; generally, it can refer to a gap in access to technology (usually the Internet), skills, usage, and information/digital literacies. Our understanding of the digital divide continues to become more nuanced and complex. The digital divide is both an international and an intra-national phenomenon, as even within developed countries, access, skills, and information/digital literacy can vary widely.63

¶37 Some of the gaps implied by the phrase “digital divide” echo other socioeconomic gaps: access, skills, usage, and literacy often vary according to race, gender, age, geographical location, and education.64 It seems likely that the digital divide may exacerbate already-existing divides, as those who have access, skills, and literacy are better positioned to participate in the digital society.65 In their study of the “disability digital divide,” in fact, María Vicente and Ana López note this very phenomenon: “even when controlling for socioeconomic characteristics such as education, income, and employment status, people with disabilities are still less likely to make internet use a part of their lives.”66 Seong-Jae Min notes that “even when everyone has equal access to media and technologies, the information gap between the haves and have-nots will not decrease because the haves typically make better use of media and technologies.”67 This implies that solving the “access” differential is not enough; skills, usage, and literacies must also be addressed.

¶38 In addition to the ongoing study of the digital divide, some scholars have posited the existence of a democratic divide. Taewoo Nam and Jennifer Stromer-Galley describe the democratic divide as “a gap in political participation between those historically enfranchised (men, whites, the better educated, and wealthier) and those not.”68 This divide persists (or is even larger) online, with the “haves” participating more in online political activity, such as communication, information seeking, and mobilization.69 Others suggest that those who are already politically active are likely to amplify their efforts with the Internet, while those who are dis-
engaged politically will continue their lack of political activity.\textsuperscript{70} Along those lines, Min found that “internet skills and political interest” were “particularly strong predictors of political internet use.”\textsuperscript{71} Here, the differentiation between mobilizing and normalizing political activity on the web is particularly apt.\textsuperscript{72} As described by these researchers, some research has found Internet use can mobilize political participation, while other research has found that political Internet activity is primarily engaged in by those who are already politically engaged, thus normalizing their activity. Hirzalla et al. note that these streams are not antithetical but “generally have different theoretical focal points” and rely on different types of empirical evidence.\textsuperscript{73}

\textsuperscript{¶39} The democratic divide, as currently described in the research, relates to the digital divide. The gaps of the digital divide are often based on different socioeconomic factors, such as gender, race, age, and education. These factors are significant when considering the digital divide in terms of access, skills, usage, and information/digital literacy. These same factors persist into the democratic divide.\textsuperscript{74}

\textsuperscript{¶40} The digital divide impacts people across a wide array of areas, including employment and job seeking, healthcare, communication with loved ones, and politics. The democratic divide, thus, can be seen as focusing on a narrow slice of the digital divide. The claim is that those who suffer from being on “the wrong side” of the digital divide are also likely to have lower online political activity, including communicating about political information, mobilizing for political purposes, communicating with representatives, and seeking political information. In turn, the democratic divide literature generally implies that this will lead to fewer opportunities and beneficial outcomes for those on “the wrong side.” In the terms of Hirzalla et al., those who suffer from the digital divide are less likely to be mobilized to become politically engaged.\textsuperscript{75}

\textsuperscript{¶41} Consider a congressperson who seeks to understand whether her constituents prefer she vote “yes” or “no” on an upcoming bill; to gauge constituent preferences, she may create an online poll on her website. Constituents with technology and Internet access, who search for political information, who are familiar with the congressperson’s website, and who are somewhat knowledgeable about the issue at hand are likely to vote in this online poll. However, constituents who lack any of the needed access, skills, usage, or literacies may well be left out of this activity. They will go unmobilized.\textsuperscript{76} Their perspectives, then, would not be heard. This is a simple yet realistic example of how the democratic divide can operate and can further exacerbate already-existing divides.

\textsuperscript{¶42} To date, the democratic divide literature has focused primarily on people’s online political activity and how such activity can be or is normalized and/or mobi-
lized. However, in this article we posit that there is another type of democratic divide, based on differential access to court documents.

¶43 As described above, court records have a unique and valuable role in holding the judicial branch of democratic governments accountable. Because of the nature of the judiciary, access to court records is the primary tool for this accountability. Counterintuitively, however, court record access has become more complex, stratified, and restricted with the move to online records. With the current PACER system, access is restricted to those who have technological access, skills, usage, and relevant literacy (specifically, knowledge of the courts in addition to basic digital literacy).

¶44 Digital divide research has established that those who are not white and who are lower in socioeconomic status or education are more likely to be on “the wrong side” of the divide. At the same time, these same individuals are also more likely to be on “the wrong side” of the judicial system, as a wealth of research through the past several decades documents; the U.S. judicial system is rife with racial disparities, for example.77 We posit that the democratic divide caused by PACER exacerbates these inequities. People who might benefit most from free and open access to court records are likely to have the most difficult time accessing them, for a host of reasons (incorporating the full spectrum of conceptualizations of the digital divide).

¶45 Applying the conceptual lens of “democratic divide” to this problem has two advantages. First, it echoes the language of the digital divide, with which researchers and citizens are already familiar; we can adapt many of the research tools and policy approaches from the digital divide to the democratic divide. There is precedent for focusing on a particular “slice” of the digital divide, such as the disability digital divide,78 yet utilizing resources conceived of for the broader digital divide.

¶46 Second, this conceptual lens focuses attention on the explicit effects on democracy and the significance of the shortcomings of this system. The digital divide nomenclature tends to emphasize overall effects on one’s life, including abilities to navigate education, the modern economy, recreation, and politics online. In contrast, the democratic divide term pulls into sharp relief the effects on democracy, on the ability of citizens to hold one branch of their democratic government accountable. Because the judiciary is arguably the least responsive branch to citizens already (because it is generally not held accountable through direct voting), any acts that limit transparency and accountability must be viewed critically. In turn, focusing on the democratic divide will help illuminate answers and approaches that may be particularly apt to the democracy aspect of this divide.

**Possible Solutions to this Democratic Divide**

¶47 There are many possible solutions to the democratic divide described here; these solutions range from the technical to the social (though we do not suggest...


78. See, e.g., Vicente & López, *supra* note 66.
that a crowd-sourced solution would be applicable here).\textsuperscript{79} We suggest that no single answer will be sufficient. Rather, a range of approaches should be undertaken to resolve this problem. To revert to solely in-person access to court records is not possible; most states and the federal government have embraced the concept of online remote access. However, this embrace is tempered with concerns about security, cost, and how much information to provide. As Arthur Sweeney notes, there is an “unrealized potential of e-government to develop citizen-government relationships”,\textsuperscript{80} we argue this holds true for the judicial branch as well.

\textsection{48} The first and most important change to be made is to the PACER system. The information architecture must be redesigned and made more user-friendly and intuitive. There are likely better ways to organize the documents to make them more accessible. For example, people have become used to searching the Internet using plain text, and allowing such flexibility on the PACER site would greatly enhance access. Metadata librarians and usability experts would be useful here.

\textsection{49} Currently, one needs an understanding of how the federal court systems are organized to locate relevant court documents, even though the average citizen (much less those who are lower in education, socioeconomic status, or formal knowledge of the judiciary) will likely not have such knowledge; for example, users may search PACER by party name but not by judge or attorney.\textsuperscript{81} Guides to creating and improving user-friendly websites and databases abound. Many government websites have become more intuitive and responsive in the past decade, and the PACER system should follow suit. In addition, revisions to PACER must address privacy and security concerns. Redaction of personal data should be consistent and strong enough to satisfy privacy advocates, but not so restrictive as to eliminate access to court documents; this will be a delicate balance to navigate, but it has been successfully done in other contexts, such as the Health Insurance Portability and Accountability Act (HIPAA) and the Freedom of Information Act (FOIA).\textsuperscript{82}

\textsection{50} At the same time, the funding model for PACER should be revisited. Two recent class actions challenge the payment scheme and whether the payment for such records is appropriate. A suit filed by a number of nonprofit organizations, \textit{National Veterans Legal Services Program v. United States}, now pending before the U.S. Court of Appeals for the Federal Circuit, alleges that any costs beyond the cost to provide records is an excess of fees under the E-Government Act of 2002.\textsuperscript{83} \textit{Fisher v. United States} alleges the overcharging of fees based on a faulty method of calculation within PACER.\textsuperscript{84} Thus, recent scrutiny concerning fees and billing as the result


\textsuperscript{81} See, e.g., Schultze, \textit{supra} note 13.


\textsuperscript{83} See \textit{C/o.sc/n.sc/s.sc/t.sc/i.sc/t.sc/i.sc/o.sc/n.sc/a.sc/l.sc A/c.sc/c.sc/o.sc/u.sc/n.sc/t.sc/a.sc/b.sc/i.sc/l.sc/i.sc/t.sc/y.sc C/e.sc/n.sc/t.sc/e.sc/r.sc}, \textit{National Veterans Legal Services Program v. United States}, https://www.theusconstitution.org/litigation/national-veterans-legal-services-program-v-united-states/\[https://perma.cc/S2RP-6K6A].

\textsuperscript{84} See, e.g., Mike Masnick, \textit{Lawsuit Against Courts Massively Overcharging for Documents Moves Forward}, \textit{Techdirt} (Dec. 8, 2016), https://www.techdirt.com/articles/20161208/00343136224/lawsuit-against-courts-massively-overcharging-documents-moves-forward.shtml [https://perma.cc/XH39-S9DE]. [Ed. note: While this article was in press, the Federal Circuit held that PACER fees were indeed being used outside of statutory authorization. Nat’l Veterans Legal Servs. Program v. United
of these suits will indeed create some uncertainty about the funding model. In addition, as noted above, the current funding model creates potential hardship for citizen information seekers from lower socioeconomic classes, leading to an explicit democratic divide. The costs of maintaining PACER and providing free, unregulated access cannot be very high in terms of the federal budget, and doing so would eliminate a very tangible barrier to accessing court records. Taken in conjunction with improving the user-friendliness of the site, this step would increase accessibility significantly.

51 A less financially burdensome option that would increase overall access to courts would be to return PACER to public libraries—allow free citizen access at these institutions. Even if PACER were not entirely redesigned, returning access to public libraries would be advantageous to all. Library and information science professionals have seen the role of information and their jobs change as the digital age has descended upon them. They represent a profession able to change, transition, and embrace the greater access to information. Traditional ideas of usage have changed due to changing technologies. In many instances, libraries are the first line of implementation in e-government enterprises, and they serve “as the main social institution ensuring access and assistance in using e-government.”

Future Research Directions

52 In this article, we describe court records and current access to them, and explicate the importance of court records for a transparent, accountable judiciary. Without such access to court records, citizens and other individuals are left in a democratic divide. With this conceptual foundation, we turn now to a brief discussion of possible research directions for the future.

53 First, it is worth stating the simple fact that we cannot prove with hard numbers that the move to online court records has decreased access. No one was tracking the number of individuals who accessed paper court records at county clerk offices and courthouses around the country, so we do not have data to compare to present online access. Nonetheless, the rate of access (and of accessibility) to court records through PACER could be quantified. This could be done through usability tests, download records, page counts, user counts, and other such data. Usability studies, in particular, could highlight flaws and weaknesses in PACER and suggest better information architecture and organization.

54 Second, case studies could be investigated, situations in which access to court records has aided in uncovering corruption or unfair dealing. If court record accessibility has some relationship to democratic functioning, perhaps it could be uncovered and traced through cases with qualitative research methods. At the same time, it is worth noting that such relationships often serve a normative function—telling us how democracies can and should work—more so than a formal causative function.

87. See, e.g., Ardia, supra note 2.
¶55 Third, the conceptual lens of a democratic divide would be most fruitfully applied internationally in cross-cultural comparisons. How do other democracies provide access to court records? How does this access, in turn, affect the transparency and accountability of the judicial system and the democratic norms in general? In some recently developed democracies, leaders have been explicit about providing transparent and meaningful access to court records.88

Conclusion

¶56 We argue that the current reliance on irregular, inconsistent online access to court records has resulted in reduced, and unequal, access. The system for accessing federal court records, PACER, is cumbersome, nonintuitive, and financially burdensome on its users. There are several reasonable solutions that could be implemented to improve access to federal court records. Doing so is important because access to court records is essential in strong democracies. We agree with Lori Anderson and Patrick Bishop that we need to measure how technology affects democracy, rather than just focusing on how much technology has been adopted.89

88. See, e.g., Gathegi, supra note 16.
Keeping Up with New Legal Titles*

Compiled by Susan Azyndar** and Susan David deMaine***

Contents

<table>
<thead>
<tr>
<th>Title</th>
<th>Reviewed by</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justice in Plain Sight: How a Small-Town Newspaper and Its Unlikely Lawyer Opened America's Courtrooms by Dan Bernstein</td>
<td>Mariana Newman</td>
<td>276</td>
</tr>
<tr>
<td>Exposure: Poisoned Water, Corporate Greed, and One Lawyer's Twenty-Year Battle Against DuPont by Robert Bilott</td>
<td>Alisha Hennen</td>
<td>278</td>
</tr>
<tr>
<td>Prosecuting the President: How Special Prosecutors Hold Presidents Accountable and Protect the Rule of Law by Andrew Coan</td>
<td>Babak Zarin</td>
<td>280</td>
</tr>
<tr>
<td>A Human Algorithm: How Artificial Intelligence Is Redefining Who We Are by Flynn Coleman</td>
<td>Aaron Retteen</td>
<td>281</td>
</tr>
<tr>
<td>The Pursuit of Happiness in the Founding Era: An Intellectual History by Carli N. Conklin</td>
<td>Matthew Neely</td>
<td>282</td>
</tr>
</tbody>
</table>

* The works reviewed in this issue were published in 2018 and 2019, with one exception from 2017 (the paperback came out in 2019). If you would like to review books for “Keeping Up with New Legal Titles,” please send an e-mail to sazyndar@nd.edu and sdemaine@indiana.edu.

** Associate Director, Kresge Law Library, The Law School, University of Notre Dame, Notre Dame, Indiana.

*** Director and Senior Lecturer in Law, Jerome Hall Law Library, Indiana University Maurer School of Law, Bloomington, Indiana.

Reviewed by Mariana Newman*

¶1 With the continued consolidation of local newspapers into the hands of large publishing companies, there is considerable nostalgia in reading a book about the publisher of a small newspaper who spent thousands of dollars, without any guarantee that it would improve the paper’s bottom line, for the sake of a principle. *Justice in Plain Sight* recounts the legal efforts of the Press-Enterprise, the local newspaper in Riverside, California, to open courtrooms during jury selection and preliminary hearings to the public and the press in the 1980s. Author Dan Bernstein is a retired Press-Enterprise reporter and columnist.

¶2 In 1980, the U.S. Supreme Court held in *Richmond Newspapers, Inc. v. Virginia*¹ that the public and the press had a qualified First Amendment right to attend criminal trials, but the Press-Enterprise’s reporters were routinely being barred from jury selection and preliminary hearings during death penalty trials in Riverside County. The newspaper’s publisher and editor believed that public access to these additional proceedings was crucial to the preservation of public confidence in the criminal justice system and, by extension, democracy. The newspaper challenged courtroom closures in two cases, known as *Press-Enterprise I*² and *Press-Enterprise II*,³ which made their way from the state courts in California to the U.S. Supreme Court.

In addition to its focus on important legal decisions with nationwide implications, *Justice in Plain Sight* embraces local Riverside County people and issues. Bernstein begins the book by detailing the three Riverside County crimes that set the controversy in motion. He also colorfully depicts his cast of characters of local attorneys, judges, reporters, and editors, bringing legal history to life.

Some of Bernstein’s organizational and content choices may seem slightly clunky to readers, though they do prove useful for achieving thoroughness. For example, in a single chapter, Bernstein reviews four Supreme Court cases that led up to the *Press-Enterprise* cases. For each of the *Press-Enterprise* cases, Bernstein also spends considerable time on the history behind, and the content of, the many briefs submitted to the Supreme Court—the *Press-Enterprise* brief, the Riverside County brief, the amicus briefs of various newspaper organizations, plus amicus briefs of other interested parties, like the attorneys of the defendants from the underlying criminal trials, the ACLU, the California attorney general, and the Riverside district attorney. Bernstein’s discussion of each brief provides an in-depth understanding of the many arguments and perspectives involved; fortunately, his investigation of the strategy behind each brief enlivens what might otherwise be seen as too much detail about a bunch of dry documents.

A Riverside lawyer, Jim Ward, who had no special expertise in constitutional law, argued both cases before the Supreme Court, somewhat to the chagrin of the experienced members of the First Amendment bar. Bernstein interviewed Ward and was given access to binders full of notes, papers, and diary entries Ward compiled during his work on the cases. These materials make for excellent documentation. Similarly, in the chapters covering oral arguments, Bernstein includes numerous excerpts from the transcripts.

Bernstein memorably quotes Ward: “Everybody thinks that attorneys have all this mystical knowledge from law school. But what we really know is how to look things up” (p.8). Legal research of the old-school variety makes two very important appearances in the book. As the *Press-Enterprise*’s legal team worked on writing its briefs in both cases, two young associates made momentous trips to the Law Library of Congress. The attorneys knew that in previous opinions related to open courtrooms, Chief Justice Warren Burger had relied on historical evidence. They wanted to find historical proof that courts were once open during jury selection and preliminary hearings. For *Press-Enterprise I*, an attorney paged through fragile reports of London’s Old Bailey trials until he found a 1660 court reporter’s comment that “the People seemed to laugh” after the defendant challenged his tenth prospective juror (p.88). As attorneys wrote their brief for *Press-Enterprise II*, another young lawyer made her own pilgrimage to the Law Library of Congress where, after a week of consulting digests and reports, she found a court reporter’s comment that the preliminary hearing during Aaron Burr’s treason trial had been open to the public; in fact, it had been so full of spectators that it had to be moved from the courtroom to a larger venue. This was an especially advantageous discovery, as Chief Justice Burger was known for his particular interest in Aaron Burr’s trial.

Fortunately, Ward and the *Press-Enterprise* were victorious in both cases, securing for the press and the public a constitutional right to attend preliminary hearings and jury selection, although exceptions remained possible under some circumstances. Likewise, Bernstein has secured for the public a look into the importance of access to information and the role of the press. Also of particular
interest is the inside look into Supreme Court procedure that *Justice in Plain Sight* provides. Bernstein relies heavily on the papers of Justice Lewis Powell, who donated his papers to Washington and Lee University, and Justices William Brennan, Harry Blackmun, Thurgood Marshall, and Byron White, who donated theirs to the Library of Congress. Bernstein makes compelling use of memos from clerks, annotated with notes from the Justices, as well as notes taken by the Justices during oral arguments and during conferences. Learning how Justice Blackmun ranked the oral argument prowess of the attorneys on his usual one-to-ten scale, how Justice Powell made note of the outcome of preliminary votes at conference, and how Justice Blackmun viewed the opinion-writing acumen of Chief Justice Burger is likely to interest those who care about the work and history of the Supreme Court.

Bernstein writes with a broad and nonlegal audience in mind, so *Justice in Plain Sight* is well worth acquiring by public libraries in addition to academic libraries of all kinds.


Reviewed by Alisha Hennen*

“‘They wanna try and keep everything hushed up. Like it’s some kind of big secret. . . . They won’t tell us what it is. They don’t wanna talk to me. Because I’m an old dumb farmer, I’m not supposed to know anything. But it’s not gonna be covered up. Because I’m gonna bring it out in the open for people to see’” (p.vii).

These were the words of Wilbur Earl Tennant, a farmer from Parkersburg, West Virginia, as he filmed his cattle dying and wasting away. Little did Earl Tennant know, as he filmed those words, that they would spark a chain of events that would uncover one of the worst cases of environmental contamination in history.

This is where the story recounted in *Exposure* starts: Tennant’s cattle were dying at an alarming rate, and no one would listen to him—not the West Virginia Division of Natural Resources, not the West Virginia Department of Environmental Protection, and especially not E. I. du Pont de Nemours & Co. (DuPont), the chemical company that owned the landfill abutting the creek that provided water to his cattle. So Tennant found attorney Rob Bilott, a corporate defense attorney specializing in environmental law and the grandson of Tennant’s friend. Bilott was not your typical environmental crusader/hero, but he took on Tennant’s case in 1998 and has been fighting DuPont ever since.

*Exposure* details how Bilott skillfully pieced together, through years of discovery battles and copious amounts of coffee, DuPont’s knowing disposal of perfluorooctanoic acid (PFOA) into the drinking water of the community surrounding its Washington Works plant in Parkersburg, including the creek on Earl Tennant’s land. PFOA, a man-made chemical used in Teflon manufacture since 1951, was unregulated at the time of DuPont’s actions but has proven to be a dangerous chemical, currently in the blood of 99% of Americans. In 2011, a panel of scientists concluded that there were probable links between PFOA exposure and kidney cancer, testicular cancer, ulcerative colitis, thyroid disease, high cholesterol, and

---

preeclampsia. Although the science is complicated, Bilott effectively translates it into comprehensible language.

¶13 *Exposure* is divided into three major parts: “The Farmer,” “The Town,” and “The World.” Each part shows how the litigation expanded as the extent of PFOA contamination was unearthed. It starts with Tennant and his determination to find out what was killing his cattle. The Tennant litigation leads to the discovery of larger contamination and a class action lawsuit on behalf of the 70,000 community members exposed to PFOA-tainted drinking water near the Washington Works plant. The book closes with tales of further individual lawsuits for PFOA-related health issues and the scientific work that showed PFOA contamination is a global problem affecting virtually everyone. The *New York Times Magazine* called Rob Bilott “DuPont’s Worst Nightmare.” Readers of *Exposure* will likely agree, but Bilott does not believe he deserves the title. As he explains, “DuPont had hoped that nobody would see the dots, or, if they did, that they wouldn’t be able to connect them. But Earl had seen the dots. Earl had connected them. I wasn’t DuPont’s worst nightmare. Earl was” (p.362).

¶14 While the story itself is compelling, it is also informative about the legal process. Bilott takes the reader through the entire progression of civil litigation, including battling for discovery, scouring over documents for months, negotiating settlement offers, and managing class action suits. The book also offers a few professional responsibility and evidence lessons. For example, *do not* email your son from your work account about the dangerous chemicals your company is releasing; if you do, those emails may end up discoverable.

¶15 If you enjoy documentaries and want to put faces to people in the story, I highly recommend watching the 2018 film *The Devil We Know: The Chemistry of a Cover-Up* directed by Stephanie Soechtig. You will hear a distressed and angry Tennant as he filmed his dying cattle. You will see an understated Bilott recount how he discovered the cover-up through documents he received during litigation. You will meet Sue and Bucky Bailey, a mother and son who have been forever affected by DuPont’s use of PFOA in Teflon. The documentary is currently available for streaming on Netflix. In addition, *Dark Waters*, a feature film about the DuPont contamination and Bilott’s fight for truth and accountability, was released in theaters in late 2019.

¶16 Overall, *Exposure* is an intriguing and easy-to-follow narrative that will have you up in arms about what might be in your drinking water. If you liked *Erin Brockovich* or *A Civil Action*, this book is a must-read. *Exposure* is highly recommended for all academic law libraries, other academic libraries, and public libraries. Firms with an environmental practice may be interested in acquiring it as well.

---


*Reviewed by Babak Zarin*

¶17 When special prosecutor Robert Mueller was first assigned to begin his investigation, many people found themselves asking what exactly a special prosecutor does and how. Andrew Coan’s *Prosecuting the President: How Special Prosecutors Hold Presidents Accountable and Protect the Rule of Law* seeks to provide answers to both legal professionals and laypersons trying to understand the role of a special prosecutor, and it does so very effectively.

¶18 *Prosecuting the President* is divided into eight chapters that form two halves. The first half, “History,” provides a historical overview of investigations by special prosecutors and the consequences of the work they did. Through Coan’s analysis of these moments—which include the work of special prosecutors during the Whiskey Ring, Teapot Dome, Watergate, and Whitewater scandals—readers learn that the reception and public treatment of a special prosecutor are complex and varied. The public’s perception of the special prosecutor’s role and the dynamics around it takes on additional importance in the book’s second half, “Law.” Here, Coan guides readers through the discussions and debates regarding the legal limits of the special prosecutor, and the relationships between special prosecutors, the president, and the general public. Coan pays particular attention to questions that are highly relevant today, such as whether a president can be charged with obstruction of justice. Coan concludes by underscoring his central theme: only the American people are able to create the cultural atmosphere necessary for a special prosecutor to be effective.

¶19 *Prosecuting the President*’s clear strength lies in the use of this structure. By creating an engaging historical account of special prosecutors in the first half of the book, Coan familiarizes audiences with this unique position in a readable way that relies on little legal jargon. This developmental approach allows readers to see how the position has changed over time, and readers will likely find the discussions of Watergate and Whitewater immensely helpful in understanding the impacts of these scandals. This approach also helps orient readers to better understand how real-world events have affected the legal discussion and subsequently affected the drafting of laws regarding special prosecutors.

¶20 Coan’s achievement is no small feat, but I must admit that those hoping to find detailed legal arguments or concrete conclusions in the second half of the book likely will be left wanting. Although Coan is meticulous in providing end-noted citations, he is writing about a legal landscape that is in many ways unsettled insofar as the work of Robert Mueller is the first major test of the powers and limitations of the special prosecutor in decades. Firm legal conclusions may be premature, but the murkiness surrounding the special prosecutor does cause the second half of *Prosecuting the President* to seem more unsettled than the first half, despite the fact that it is well organized and clearly written.

Coan’s work in creating a highly readable book that comprehensively portrays the historical, legal, and political atmosphere in which special prosecutors operate is a feat well worth praising. That the book was published in 2019 and avoids making extreme claims makes it additionally praiseworthy, as it addresses a pressing need without a display of partisanship. Notably, *Prosecuting the President* is the first major work focusing specifically on special prosecutors to have been written in decades, and among few such works ever. *Prosecuting the President* is a welcome and important addition to law, college, and public libraries.


Reviewed by Aaron Retteen*

When are the moments to watch for, and what actions must be taken, to prevent a dystopian future of super-intelligent machines controlling human activity? Can we instead ensure a utopian future of humans working alongside machines for the benefit of human society? *A Human Algorithm* provides readers with a provocative overview of the state of technology today and challenges us to work together to avoid foreseeable pitfalls while the window for human influence on this matter remains open.

The book proceeds chronologically, taking readers on an enjoyable journey highlighting historical milestones in the development of modern computing technology, while underscoring the many lessons we can draw upon to ensure future technological innovations do not negatively disrupt human life. As an avid reader of the history and potential future of technology, I was impressed by Coleman’s research: this book is a great read for technology experts and novices alike. Unlike some publications on this topic, this book does not require any specialized knowledge in artificial intelligence (AI) and has broad appeal. Coleman’s writing is neither intimidating nor technically cumbersome, and the arc of the book focuses on philosophical questions surrounding the development of AI that can engage even the most technologically inept among us. Included throughout the book are helpful figures and diagrams that visually depict larger points, and long timelines that make understanding the author’s perspective a bit easier.

To help readers break down the many important questions that have arisen alongside the onset of AI, Coleman analogizes to prior technological developments through a lens of potential costs and benefits. Many people will remember learning about the Renaissance or the Industrial Revolution, but this book shines light on how specific developments in information technology were catalysts for greater societal changes during those eras. This helpful and effective approach provides readers with familiar context when thinking about technology that may otherwise be unfamiliar. The author also encourages readers to break away from thinking about intelligence from a human-centric perspective and points to examples, such as how octopi use a different (nonhuman) intelligence system to simultaneously and independently perform tasks, that demonstrate how we can look beyond our

* © Aaron Retteen, 2020. Instructional Assistant Professor of Law and Digital Services & Repository Librarian, Dee J. Kelly Law Library, Texas A&M University School of Law, Fort Worth, Texas.
own nature when designing artificially intelligent technologies. While AI pioneers, including individuals, large businesses, and even governments, work today on the innovations of tomorrow, this book encourages us to stop and think about what it means to be “intelligent,” and whether humans should be—or, indeed, whether humans are even capable of—guiding artificially intelligent creations to have a sense of ethics or morality.

¶25 My favorite aspect of the book is that the author does not shy away from underlining how difficult it will be for human society to come together in the ways needed to avoid some of the most pernicious potential outcomes of AI technology development. Our society is already, in some important ways, on the wrong track: racial and gender diversity in computing and AI is low, and the barriers to entry are high, both in terms of the specialized knowledge and the capital investment needed to create new forms of artificially intelligent technology. Coleman does not ignore these facts; instead, she embraces them as evidence for the argument that we are at a point ready for reflection.

¶26 The author impressively weaves the book’s chapters in a way that combines potential individual impacts, such as data privacy and job loss due to automation, with large-scale societal impacts, such as AI use by militaries and AI technology reinforcing social biases and inequality, leaving readers with the conclusion that we must take control of our future before it is too late. By the end of the book, there is no escaping the fact that developments in AI technology will be made alongside social and political developments, and that the former is greatly influenced by the shape of the latter. Although the stakes are high and the possibility of a dystopian AI future feels real, the author does not wallow in despair or defeat. The “human algorithm” that has the potential to negatively shape the AI of the future has a similar potential to be beneficial to individuals and human society. The key question that the author leaves open to the reader is whether human society is ready to embrace our differences, learn from each other and our past, and transition into a new information age. If we are, we will be capable of transforming our society along with technology to create a better world.

¶27 For anyone looking for an easy read about the history and potential future of artificially intelligent technology, A Human Algorithm is a great resource that will likely change the way you think about technology, politics, and humanity as a whole. I highly recommend it for academic law libraries, as well as other academic libraries and public libraries.


Reviewed by Matthew Neely*

¶28 The Pursuit of Happiness in the Founding Era: An Intellectual History illuminates the enigma of the phrase “the pursuit of happiness.” Part of a series entitled Studies in Constitutional Democracy, this book investigates the meaning of “the pursuit of happiness” in the context of the origin and development of American constitutional and democratic traditions. Legal historian Carli Conklin looks to the
time before the Declaration of Independence to examine the idea of happiness as it was conceived of in 18th century England and colonial America. She also explores the phrase’s application in the early history of the new republic.

§29 *The Pursuit of Happiness in the Founding Era* begins by examining in detail the philosophical and theological underpinnings of Blackstone’s definition of the phrase in his *Commentaries on the Laws of England*. Blackstone concluded that the pursuit of happiness was to seek to live per the two “books” of divine revelation—Holy Scripture and nature (i.e., natural law written into creation by God)—by applying human reason and the individual conscience to the study of these two sources of truth (p.30). Conklin demonstrates that this erudite understanding did not originate with Blackstone but rather reflects a classical idea influenced by Anglican latitudinarian theology and the Scottish Enlightenment. Blackstone and others then applied the phrase to legal considerations, in turn shaping the common law by way of their academic examination and teaching. Conklin’s investigation highlights the convergence of these four different influences—classical literature, Scottish Enlightenment philosophy, Anglican theology, and English legal history—in Thomas Jefferson’s “pursuit of happiness” in the Declaration of Independence. Conklin believes Jefferson and other founders understood the phrase “in the natural law sense of man’s right to live a rightly ordered or virtuous life in accordance with the law of nature as it pertains to man” (p.96).

§30 Conklin goes on to examine the dual nature of the pursuit of happiness as a private right (the right to individually pursue virtue as dictated by individual conscience and reason) and as a public duty (the duty to live according to virtue within the community and to improve the common law). Together, these two aspects of happiness form the basis of a “science” of jurisprudence (p.113). Blackstone believed that a “scientific” approach to jurisprudence would enable lawyers, judges, jurors, and members of Parliament to conform the law to the principles of reason and the natural law, thus promoting human flourishing and fulfilling their public duty. Blackstone and Jefferson both used this scientific jurisprudence to argue for extensive revisions to the criminal law of England and the early United States, respectively. Notably, Blackstone argued that criminal sentences should not only punish but also rehabilitate offenders. Jefferson argued for limiting the death penalty to treason and murder. The notion of the pursuit of happiness made such reforms possible even though Blackstone and Jefferson were philosophically and politically hostile to one another.

§31 Conklin’s book is well organized and includes six appendices reprinting primary sources, endnotes, and an index for both the main text and the appendices. The endnotes are detailed and contain extensive explanatory material. I found the use of endnotes rather than footnotes to be the book’s main weakness in terms of format, as I had to continuously flip from the text to the endnotes to check whether

5. *Eudaimonia*, or the classical Greek understanding of human flourishing by living according to the dictates of virtue.

6. Latitudinarianism was a school of thought that embraced the notion that adherence to particular liturgies, church forms, and doctrinal minutiae were not important to God; rather, following truth determined through human reason, in freedom, was the means to salvation.

7. Blackstone argued that reform should occur only through acts of Parliament or the courts. Jefferson believed that the pursuit of happiness dictated that individuals overthrow a tyrannical government.
a note was merely a citation or something more substantive. That said, the end-notes exhibit a breadth of resources that provide researchers many avenues to continue their research.

¶32 An engaging read, this book is an excellent addition to any academic law library. Researchers will find this book useful not only as a historical treatment of “the pursuit of happiness,” but also as a springboard for further research. Law firm and government law libraries will find that this work is too theoretical for practical use, although historically savvy attorneys might find it an interesting read.


Reviewed by Jennifer Mart-Rice*

¶33 Assessment: this buzzword, always lingering about in libraries and related organizations, can cause both excitement and controversy. Assessment can have a variety of meanings and corresponding reactions, depending upon the individual. However, I think everyone can agree that we all want to improve our libraries, make our collections more responsive, and communicate the importance and relevance of what we do within the library or organizational setting. Assessment Strategies in Technical Services works hard to help technical services librarians and staff do just that.

¶34 Assessment Strategies in Technical Services begins with an introduction to the various assessments possible in a technical services department, which are examined in later chapters. It also provides context for the reader by comparing “traditional metrics and assessment” with “contemporary metrics and assessment” (pp.4–11). This introduction is thought-provoking, and I found myself comparing how my own departmental assessments fit into either the traditional or contemporary categories and how our recent integrated library system (ILS) upgrade could help with more nuanced assessments. Some assessments that were not possible in previous systems without purchasing add-on software or programs, e.g., user statistics for the library catalog, a breakdown of expenditures for each individual vendor or resource type, and compiling statics for U.S. News or ACRL, are now becoming easier and/or possible to perform. After the introductory chapters, the focus of the book shifts to individual areas in which assessment can be performed and possible avenues to pursue. Some of the topics covered in these chapters include e-resources, serials, cataloging, and metadata.

¶35 A bonus feature is the various case studies found in each of the chapters. These examples applied to real-world libraries are insightful; they show how different assessments have been performed and what institutions have learned from the assessments themselves, as well as the assessment process. Additionally, each chapter contains numerous citations to a variety of resources that go into more depth on particular assessments and/or discussion points. Some chapters also have a “Resource” section with additional case studies, templates, and datasets. The two chapters that law librarians may find most instructive concern holistic assessment (chapter 2) and serials (chapter 4). These two chapters are particularly interesting.

and could be very helpful to law libraries as they create strategies for evaluating their most financially cumbersome resources.

¶36 While I found *Assessment Strategies in Technical Services* to be quite informative and stimulating overall, I did find some of the chapters to be technical and dense. This could be overwhelming for a new director or technical services librarian who is embarking on his or her first assessment, or for anyone transitioning from “traditional” assessment measures to more “contemporary” assessment measures. However, this anxiety could be easily managed by choosing one resource type to assess at a time, reading and rereading the relevant chapter and its corresponding references, carefully planning a timeline and steps per the guidelines, and taking each of the assessments one step at a time. As the authors of the book continually point out, each library is different, and although librarians can look to another library for guidance or as an example, their libraries likely have different missions and goals, smaller or larger staffs, varying resource types, and different audiences than the example library. Library professionals need to be flexible and open to alternatives in their assessment planning.

¶37 This book covers all of the technical services bases. It can assist the reader in evaluating their existing collections, creating reports to justify expenditures or positions, and examining workflows that may have become stagnant as an ILS has evolved or been upgraded to a next-gen ILS environment. While this particular title may not fit into a defined legal subject area that is outlined in a law library’s collection development policy, it is a useful tool to have in the library’s arsenal. I recommend the purchase of *Assessment Strategies in Technical Services* for the library’s professional development or internal collection. In fact, I purchased multiple copies for my department so that we can plan a more thorough assessment of some of our collections over the next few years.


Reviewed by Jennifer E. Chapman*

¶38 Historian Eric Foner has a deep knowledge of the Reconstruction Era, as shown through his extensive writings on this pivotal time in American history. In his latest book, *The Second Founding: How the Civil War and Reconstruction Remade the Constitution*, Foner takes the reader on a constitutional history journey beginning at the end of the Civil War, through passage of the Thirteenth, Fourteenth, and Fifteenth Amendments, and ending with some of the Supreme Court cases that reshaped these amendments. Through this journey, Foner shows how these amendments altered political, societal, and legal systems—initiating a “second founding.” He emphasizes that the historical and social contexts in which the Reconstruction Amendments were ratified are vitally important to understanding the impact the amendments have had on the present day. In his preface, Foner acknowledges that the Reconstruction period is “unfamiliar to many” but, he states, “Reconstruction remains part of our lives, or to put it another way, key issues confronting American society today are in some ways Reconstruction questions” (p.xxi).

¶39 Foner sets the historical context and then leads readers on the bumpy paths each amendment took to ratification, dedicating a chapter to each amendment. He follows with a chapter reviewing court cases that reshaped and limited the amendments’ reaches. Throughout the book, Foner uses language from the post–Civil War era to paint a complete picture of the historical moment, quoting politicians and citing primary sources, and effectively capturing the emotions and thoughts driving political and public opinion of the day.

¶40 In his discussion of the Thirteenth Amendment, Foner acknowledges that “[t]he deceptively straightforward language of the Thirteenth Amendment raised profound questions about American society,” specifically what the “meaning of freedom” was to newly freed African Americans (pp.40–41). The Thirteenth Amendment “initiated [a] . . . redefinition of federalism” and made possible the Fourteenth Amendment as we know it today (p.32). Foner emphasizes this historical significance by ending his examination of the Thirteenth Amendment with a quote from Massachusetts Senator Charles Sumner: “Liberty has been won. The battle for equality is still pending” (p.54).

¶41 In his treatment of the Fourteenth Amendment and the significant debates and compromises that led to its ratification, Foner lays out the many political and societal issues that shaped the amendment. Specifically, he highlights that the amendment was born out of negotiations and political wrangling that led to a profound constitutional shift, but that also left both parties disappointed. Representative Thaddeus Stevens captured a Republican sentiment that the amendment was imperfect but it was best to “leave it to be perfected by better men in better times” (p.87). Foner later examines the difficulties “better men in better times” have had in interpreting the amendment, especially the amendment’s terminological ambiguity.

¶42 The Fifteenth Amendment did what the others did not—it settled who had the right to vote. Although the amendment consists of only two sentences, Foner’s thorough historical review of its passage and the subsequent violence in implementing voting rights emphasizes the amendment’s significance. While the Fifteenth Amendment was the final Reconstruction Amendment, the second founding did not end with its ratification. As Foner goes on to recount, the Supreme Court reshaped these amendments. He is critical of the Court, stating that it “played a crucial role in the long retreat from the ideals of Reconstruction” (p.127). Law students and lawyers will recognize many of the cases Foner reviews, including the *Slaughterhouse Cases* and *Plessy v. Ferguson*, but by using the lens of a historian, Foner expands the traditional constitutional law class teachings.

¶43 Though less than ten pages long, Foner’s epilogue is the critical culmination of his contention that the ideals of the Reconstruction Amendments have not yet been met, and much of the blame resides with the Supreme Court. Like a lawyer building the theory of a case through factual evidence to support his final argument, Foner uses the first four chapters of his book to support his closing criticism of the Supreme Court. Specifically, he blames the Court for reshaping the Reconstruction Amendments counter to the intent indicated by the historical record.

¶44 In examining the politics of the Reconstruction period and the history of these three vital amendments, Foner reminds readers that some of our present-day political disputes are not new. We are not living in a time devoid of influence from the historical record; contemporary political and social issues are all part of the
long continuum of history. This is not to suggest we are helpless to the push of history. Rather, we must remember law is not stagnant, and understanding legal history is a significant part of understanding present-day law. As Foner says, “[O]ur understanding of the Reconstruction Amendments will forever be a work in progress” (p.176).

¶45 Foner has the enviable ability to synthesize vast amounts of historical information and compose an engaging narrative. The Second Founding is best read as a history book that gives deeper meaning to the Reconstruction Amendments, rather than as a constitutional law book. It is recommended for academic law libraries and other college and university libraries.


Reviewed by Katie Ott*

¶46 It is not often that a novel-length gripe about the American legal system proves both humorous and readable; yet Bruce Gibney, a lawyer turned venture capitalist, accomplishes just that in his new book, The Nonsense Factory: The Making and Breaking of the American Legal System. In an analysis that examines the systemic problems plaguing the legal system, Congress’s obsession with reelection at the expense of passing meaningful reforms, and the imbalance of power among the three traditional branches of government and the federal regulatory state, Gibney uses subtle humor and commonsense logic to deliver a grim but trenchant diagnosis of what’s ailing the American legal system.

¶47 Beginning with legal education in the United States, Gibney snidely pokes fun at the American process for educating new lawyers, stating that “American law schools have settled for the worst of all worlds—an abstracted research faculty presiding over an outmoded trade school. The results have not been good” (p.46). According to Gibney, the inadequate, ill-trained lawyers emerging from law schools are only the start of the problems plaguing the American legal system.

¶48 Gibney’s view of the present state of Congress is similarly grim. He contends that rather than legislating for positive change, members of Congress spend a grotesquely inordinate amount of time, money, and human capital on just getting reelected. In fact, Gibney writes, campaigning has become “so intense and expensive that ex-Representative Tim Roemer suggested it censuses between 30 and 50 percent of a member’s calendar, and other Congressional insiders concur” (p.93). When representatives spend so much time and energy on campaigning and elections, they leave themselves little time for actual legislating.

¶49 Gibney’s deepest anxiety stems from the imbalance of power among the three traditional branches of government and the regulatory bodies that have become inextricably woven into the American legal system. Agencies “constitute most of the federal government” yet, alarmingly, “we cannot say precisely what all these agencies do” (p.111). Gibney equates agencies to bureaucratic platypuses—creatures with such a bizarre taxonomy that they do not fit into a single established branch of power (an agency is neither judicial, executive, nor legislative). While

Gibney acknowledges the Administrative Procedure Act technically places agencies under the executive branch, he worries that many agencies remain independent of the President’s influence, making them a quasi-fourth branch of government. Gibney bemoans that “the legal status of bureaucracies has always been uncertain, and agencies persist only through acts of willful blindness, legal contortionism, inertia, and expedience” (p.130)—a harrowing observation about an aspect of the U.S. government that holds an impressive amount of power but has startling little oversight.

¶50 While *The Nonsense Factory* excels in detailing the many problems with the American legal system (later chapters cover prosecutors, public defenders, and prisons), Gibney leaves readers with little guidance as to how we might solve them. Rather than offer concrete, achievable solutions, Gibney falls back on frustrated gripes about the self-serving practices of government officials. For example, he complains that judges “creat[e] deliberate ambiguity over the administrative state [to] amplify judicial influence over the entire government. This is probably as some judges prefer, for judges have long seen themselves as not only the center of the legal universe, but the brightest stars within it” (p.131). Gibney’s fuming over the self-serving practices of the judiciary feels more petulant than constructive for readers hoping to enact change and improvement within the legal system. But perhaps the feelings of frustration and failure are intentional? Gibney notes that “the law is bad, but not *that* bad, not yet” (p.379). So things can, and very possibly will, get worse.

¶51 Gibney’s take on the American legal system is both well argued and readable, and it would make an excellent addition to an academic library’s collection. The deep research Gibney committed to the cause (see the 99 pages of notes at the end of the book) provides a legitimacy to the mounting sense of doom that grows with each page of *The Nonsense Factory*. Humor, dispersed throughout the chapters, largely makes the read enjoyable (if the reader can overlook Gibney’s incessantly gloomy outlook on the legal system as a whole), even for those lawyers or academics who may not find Gibney’s criticisms particularly unique. Gibney makes steady use of both current and historical examples to back up his claims, which gives *The Nonsense Factory* a historical yet timely tone. As a result, Gibney’s address leaves the reader with a clear understanding of the flaws in the United States’ legal structure. Of course, it remains to be seen whether his clarification of existing problems will be enough to stimulate serious change.


*Reviewed by Andrew W. Lang*

¶52 With the advent of the web and social media sites, people have become increasingly aware of the ephemerality of our media sources. This general awareness of media’s instability is less surprising to librarians and archivists, who have been waging war against the forces of deterioration, decay, and obsolescence for thousands of years. While some losses in the cultural record resonate at a massive
scale—Shakespeare’s Cardenio, anyone?—it is the cumulative effect of the smaller, day-to-day losses that create historical amnesia. Authors Kathleen A. Hansen and Nora Paul take up this theme in *Future-Proofing the News: Preserving the First Draft of History*, focusing specifically on news media.

¶53 As the authors discuss in their introductory chapters, in a marketplace where producers of news media are preoccupied entirely with the present, they tend to devote little thought or energy to yesterday’s news. Yet daily news sources are the “first draft of history,” and although understanding of current events is often refined through distance and subsequent information, immediate reactions to events provide one of our best tools for assessing historical attitudes. Historical news coverage can also serve as a corrective to subsequent narratives, which have the disingenuous tendency to portray events as inevitably leading to a forgone conclusion, obscuring the messy reality of living through history.

¶54 After laying this groundwork, the authors devote each subsequent chapter to a different media format, arranged in chronological order so that, when read straight through, the work provides a fascinating and concise history of mass media technology. The major categories discussed are print newspapers, visual news, film newsreels, radio news, broadcast television news, and digital formats.

¶55 Each chapter in turn conforms to the same structure: a descriptive account of how the format developed in the context of news coverage, why materials in that format were lost, how some materials were preserved, and a description of the ongoing preservation challenges. The narrative is illustrated with case studies, highlighting the heroic efforts, natural disasters, serendipitous flukes, and fatefully short-sighted decisions that have shaped the current landscape. For example, when the Library of Congress decided to convert its print news collection to microfilm in the 1940s, it cut the pages from their binding to get a better image, but the cutting rendered it impossible to rebind the pages post-filming, so they were discarded. Many other libraries followed suit, converting “tens of millions of pages of newspaper pages into microfilm” (p.34). You can almost hear the millions of librarians suddenly cry out in terror. It was not until a few decades later, however, that the full gravity of these decisions became apparent when librarians began discovering that “vinegar syndrome” had slowly been eating away at these microfilm collections; in 1991, a librarian at the University of Florida discovered that three-quarters of the library’s master negatives had deteriorated, including the only back files of many Florida newspapers (p.35).

¶56 For broadcast media such as radio and television, much of the early news content was lost because it was never recorded in the first place; the famous recording of the Hindenburg disaster was captured because it was coincidentally part of an experiment in recording for time-shifted broadcasting. Similarly, in 2016 when the *New York Times* wanted to commemorate the twentieth anniversary of its online platform’s launch, it had no screen captures from the release date; the closest it could get—through the Wayback Machine,8 no less—was from nearly 11 months post-launch.

¶57 After tracing these patterns across each new media format, the authors highlight three overarching themes:

For the most part, the actual creators of the content rarely concerned themselves with preserving the news product; when news was preserved, the material either used to create it or to store it often was insufficiently stable to ensure future accessibility; but for the efforts of a wide array of organizations, foundations, libraries, memory institutions, and individual collectors, the body of news content that has been created during the past three centuries would be lost (p.215).

There is something tragic about how this cycle repeats every time a new media source is added to the ecosystem, and it is my impression that the authors hope drawing attention to this cycle may prompt people to be more proactive when the next big technology rolls around.

This book presents an excellent introduction to the last three centuries of media history broadly, but it will be most valuable to those investigating the subject of multimedia conservation for the first time. At under 250 pages, the focus is on breadth rather than depth, though every chapter is well sourced with endnotes, and the authors append a helpful list of suggestions for further reading.

For law librarians, this work may seem niche, but it is easy to see parallels to the challenges we face in our own information ecosystem. As our collections increasingly move online, we face questions about their longevity, stability, reliability, and accessibility, always with the understanding that losses and lapses in preservation can have dire consequences. There is no silver bullet solution to the issue of preservation, but a proactive concern seems like a great place to start. This book is recommended for academic law libraries, college and university libraries, and larger public libraries—any library with an interest in the cultural importance of information preservation.


Reviewed by Gilda Chiu-Ousland

In an environment where increased scrutiny has been placed on libraries to show their value within the institutions that house them, strategic planning has become a useful instrument to convey how the library’s goals fit and uphold the mission of the larger institution. It also provides libraries with opportunities to evaluate their organizational structure, processes, and use of resources in greater depth as a means toward improvement. However, embarking on the strategic planning process is often easier said than done. Strategic Planning for Academic Libraries attempts to provide a checklist of steps, questions, and tips to make the strategic planning process work for any librarian.

The book is divided into seven chapters. Each chapter, with the exception of the last, focuses on one of six steps the authors have identified as important to the strategic planning process, from preplanning to assessment. These chapters follow a similar structure: conceptual overview of the specific planning stage, plus common issues and challenges associated with the stage; a real-world example from the authors’ library, the University of Utah’s J. Willard Marriott Library; and

reflections garnered after the process at the authors’ library, aptly named “Lessons Learned.” The concluding chapter acts as a wrap-up of everything discussed in previous chapters with a focus on highlighting the essential components of a smooth and successful strategic planning process. The book also provides a wealth of appendices related to the strategic planning process at the J. Willard Marriott Library. Some examples of the included appendices are the library’s former and current strategic plans, the survey used to solicit feedback, the final report and recommendations resulting from the process, and the focus group questions used to collect data from stakeholders.

¶62 The style of Strategic Planning for Academic Libraries is best described as that of a well-organized and functional guidebook. Concepts are introduced and succinctly explained through citations from current strategic planning literature and the authors’ own experiences. The authors include various subsets of questions throughout the book to direct readers from conceptual to practical application, making the concepts more accessible and usable for those not as familiar with them. The strategic planning process is outlined in a methodical way, with particular attention paid to why certain actions were taken by the authors and how those actions were executed. Concrete examples from the authors’ strategic planning process explicate each stage of the process.

¶63 One of the strengths of this book is how it excels in effectively demonstrating the implementation of the theoretical framework the authors created and applied to their strategic planning process at the J. Willard Marriott Library. They offer a detailed account of each step in the process, what worked and, most important, what did not and how they would improve their strategic planning methods further.

¶64 Strategic Planning for Academic Libraries is a valuable resource for academic law librarians who are interested in or planning on creating a strategic plan for their organization. It offers a clear and pragmatic approach substantiated with real-world application through the case study. The book is also recommended for other law librarians who want a general overview of the different elements and considerations involved in a strategic planning process. Those looking for a deep dive into strategic planning theory or practices should seek alternatives to this book. Strategic Planning for Academic Libraries is very much a how-to guide rather than an exhaustive look at strategic planning.


Reviewed by Pamela G. Smith*

¶65 Legal Research Demystified is authored by Eric P. Voigt, a legal research and writing professor at Faulkner University’s Thomas Goode Jones School of Law. The book is designed to serve as a legal research textbook for first-year law students. Voigt’s problem-typing approach breaks the legal research process down into steps, providing a logical method that students can grasp and apply in their future work.

Like many books of its kind, *Legal Research Demystified* opens with chapters devoted to the U.S. legal system, binding and nonbinding legal authorities, and the publication process, as well as the anatomy of cases and statutes. While most other legal research texts then proceed to devote single chapters to researching secondary sources, case law, and statutes, *Legal Research Demystified* sets a different course. Voigt divides legal research problems into two basic areas: case-based common law questions and statutory questions. In researching common law, Voigt walks students through eight steps: (1) creating a research plan; (2) using secondary sources; (3) researching cases by topic; (4) using a citator to check the validity of cases and to find additional cases; (5) finding cases using a keyword search; (6) expanding case law research within the jurisdiction; (7) researching cases outside the jurisdiction; and (8) using a citator to reconfirm the validity of each case. For issues that begin with a statute, Voigt uses nine similar steps that include locating, validating, and reading statutes, regulations, and interpretive cases. He adds legislative history research as a step as well, which many instructors will appreciate. With these step-by-step approaches and clear examples to differentiate between researching common law problems and statutory law problems, Voigt addresses a challenge students often face—understanding the interplay between case law and statutory law.

Voigt covers print as well as electronic research, including a section entitled “Print Research: Four Reasons to Know It.” Instructors are likely to appreciate this section because it emphasizes what most of us are telling our students about why they still need to know how to research in print. In terms of electronic research, while Voigt gives a nod to other electronic databases, his primary focus is on how to conduct research using Westlaw Edge and Lexis Advance.

Voigt’s text is clear and easy to understand, and it is supported by illustrations, charts, and examples that add to his goal of making legal research easier for students to understand. Many of these are captured directly from the various print and electronic resources that are discussed in the text. Others, such as a table of “Binding Versus Nonbinding Authority” and the “Research Action Plan,” are variations of materials that many of us are already using in the classroom, but students will find it helpful to have this information included in their textbook.

Multiple-choice and true/false questions are also included with many of the chapters, allowing students to practice and test their knowledge of the material. Unfortunately, though, answers to the questions are not provided in the book itself. Instead, answers are provided through Core Knowledge, an interactive website operated by Carolina Academic Press. Students who purchase a new copy of the book receive a code to access Core Knowledge. Students who purchase a used copy have the option of paying for access to Core Knowledge. It is unclear from the Carolina Academic Press website exactly how much a student who purchases a used book has to pay for this access. Carolina Academic Press advertises Core Knowledge as containing interactive research exercises the professor may use to flip the classroom. Based on this description, I was excited to test Core Knowledge, but after logging in using the code I received with the book, I discovered that the only content, interactive or otherwise, available for *Legal Research Demystified* is the same questions that appear in the book. These questions are laid out in an online quiz format in which students answer questions and then receive feedback.
based on their answers. While this quiz is certainly useful, I would like to see more robust content such as videos and tutorials added to Core Knowledge.

In sum, *Legal Research Demystified* is recommended for academic law library collections. Also, for anyone who teaches legal research, it is worth considering as a textbook based on its clear, step-by-step approach. However, additional content to Core Knowledge would make *Legal Research Demystified* even more effective. Hopefully, this content can be enhanced in the future.