Legal Research in the Computer Age: A Paradigm Shift?*

Carol M. Bast** and Ransford C. Pyle***

The digest system represents the present paradigm of legal research. The authors contend that there will be a gradual shift to a new legal research paradigm as a greater percentage of research is accomplished on the computer. The new legal research paradigm will be the computer code used in computer-assisted legal research.

¶1 A coherent legal system had developed in the United States by the end of the nineteenth century, along with a stable federal and state appellate court structure and the case method of law school instruction. In the realm of legal research, the same period saw the development of the key number system, digests, citators, and annotated reporters.[1] Of these, digests especially bear the marks of the time in which they were created. Nonetheless, the digest system still represents the present paradigm of legal research.

¶2 But a revolution in legal research is taking place right now because of technological change. Given its capacities for storage, organization, and retrieval, the future of legal research is inevitably linked to the computer. With computers, researchers can formulate their own word searches rather than rely entirely on the predetermined indexing of a digest. The researcher can use full-text retrieval to locate significant unusual terms and crucial information that may have been omitted from the print index. Similarly, it can also be used to find recently coined terms that don't appear in the digest thesaurus and cases too recent to have been included in the print digest.

¶3 Whether the organization of printed materials in the modern law library molded legal concepts or whether the two simply are intertwined is unclear.[2] What

---

** Associate Professor, Department of Criminal Justice and Legal Studies, University of Central Florida, Orlando, Florida.

*** Associate Professor, Department of Criminal Justice and Legal Studies, University of Central Florida, Orlando, Florida.

1. Carol M. Bast, *What's in a Name? Shepard's Citators*, 14 J. PARALEGAL EDUC. & PRAC. 41, 41–42 (1998).
2. A number of authors have argued that the organization of the print law library molded the development of the law. *See, e.g.*, Robert C. Berring, *Legal Research and Legal Concepts: Where Form Molds Substance*, 75 CAL. L. REV. 15, 15, 26 (1987); Thomas J. Young Jr., *A Look at American Law Reporting in the 19th Century*, 68 LAW. LIBR. J. 294, 306 (1974). One author sees the two as inter-

is clear is that the computer will have an increasing effect on the legal system; several authors have speculated that computer-assisted legal research (CALR) will change legal concepts.[3] Gradually, we will be shifting to a new legal research paradigm as a greater percentage of research is accomplished on the computer. The new legal research paradigm will be the computer code used in CALR.

¶4 This article explores the effect CALR will have on legal thinking. The first two sections introduce the concept of a paradigm shift and discuss the characteristics and use of digests. Following sections describe print and CALR indexing and CALR research. Finally, the article discusses the legal research process and legal research of the future.

A Paradigm Shift

¶5 In his book, *The Structure of Scientific Revolutions*,[4] Thomas Kuhn describes progress in a scientific field. Scientific work in a field is performed against the background of a foundational concept or paradigm. Progress in the field is cumulative, with the paradigm as the basis for research. A scientific revolution, or paradigm shift, occurs when a new paradigm is accepted by enough scientists in the field that the new paradigm supplants the old one.

¶6 One might view the digest system as having been the legal research paradigm for the past century. Legal research is at the brink of a paradigm shift from the digest system to a system established by CALR.

¶7 The replicating characteristic of the law has been noted by a number of authors.[5] In his book, *The Selfish Gene*, Richard Dawkins coined the word "meme." A meme is the nonbiological equivalent to a gene. A meme is a concept that replicates by being transmitted from person to person.[6] Successful long-term

---

twined. "While legal indexes may have influenced the conceptual coherence of the law, they are as much a product as a progenitor of that conceptual structure." Nazareth A. M. Pantaloni III, *Legal Databases, Legal Epistemology, and the Legal Order*, 86 LAW LIBR. J. 679, 699 (1994). Another author refuses to enter the chicken and the egg debate, but recognizes that the two are "inextricably intertwined." Barbara Bintliff, *From Creativity to Computerese: Thinking Like a Lawyer in the Computer Age*, 88 LAW. LIBR. J. 338, 344 (1996).

3. *See* Robert C. Berring, *Full-Text Databases and Legal Research: Backing into the Future*, 1 HIGH TECH. L.J. 27, 29 (1986) (citation omitted) (suggesting that more work needs to be done on the relationship between the form and substance of the law); Young, *supra* note 2, at 306 (suggesting that a study be done of the effect of legal and government databases on the law). One author claims that technology may have a negative impact on the attorney's ability to engage in legal reasoning. Molly Warner Lien, *Technocentrism and the Soul of the Common Law Lawyer*, 48 AM. U. L. REV. 85, 85 (1998). She suggests that if attorneys are cognizant of this potentially negative impact, they can use technology wisely. *Id*.
4. THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (3d ed. 1996).
5. *E.g.,* Michael S. Fried, *The Evolution of Legal Concepts: The Memetic Perspective,* 39 JURIMETRICS J. 291, 292 (1999); Richard Delgado & Jean Stefancic, *Why Do We Tell the Same Stories? Law Reform, Critical Librarianship, and the Triple Helix Dilemma,* 42 STAN. L. REV. 207, 208 (1989).
6. RICHARD DAWKINS, THE SELFISH GENE 192 (new ed. 1989).

replication depends on "longevity, fecundity, and copying-fidelity."[7] Digests are a meme vehicle because their conservative organizational structure has facilitated the replication of legal concepts in successive generations of case law. In addition, the digest classification scheme is learned by successive generations of law school students; this comprehensive classification of the law underlies the attorney's approach to the law.[8]

¶8 The digest scheme is a product of the era of the legal formalists.[9] "In the Langdellian era . . . [p]rofessors at the newly invigorated university law schools could claim to take a place in their laboratories—law libraries—searching out the imperishable truths of legal science just as their colleagues in chemistry, biology, or physics were doing for their respective sciences."[10] What emerged under the influence of Langdell was a classification scheme. "[L]aws are commands, legal decisions can be deduced logically from predetermined rules . . . ."[11] The law was divided into broad categories, such as contracts and torts. Within each category were legal principles. In Langdell's scheme, "[a] few basic top-level categories and principles formed a conceptually ordered system above a large number of bottom-level rules. The rules themselves were, ideally, the holdings of established precedents, which upon analysis could be seen to be derivable from the principles."[12]

¶9 The first-year courses Langdell established at the Harvard Law School track the digest classification scheme. The major digest classifications—property, contracts, torts, and crimes—are the subject matter of introductory law school courses. Individual digest topics are the subject matter of other law school courses. Langdell's model of legal education spread from Harvard to all other law schools.[13] Langdell's influence was also felt in the print sources first published at the end of the nineteenth century. "Given Langdell's stress on categorizing the law into inflexible subject areas, and the focus on law as finding primary sources, the legal information system played an especially dominant role."[14]

¶10 In one article, the authors compare the digest system to the double helix of DNA, "enabl[ing] the current system to replicate itself endlessly, easily, and

---

7. *Id*. at 194.
8. In his recent article considering the impact of memetics on the law, Michael Fried noted that "[r]ules of law, precedents, and legal doctrines can be treated as memes competing to appear in legal reporters." Fried, *supra* note 5, at 307. Although not mentioned by Fried, we consider the digest system a meme vehicle.
9. Steven M. Barkan*, Deconstructing Legal Research: A Law Librarian's Commentary on Critical Legal Studies,* 79 LAW. LIBR. J. 617, 636 (1987).
10. M.H. Hoeflich, *Law & Geometry: Legal Science from Leibniz to Langdell*, 30 AM. J. LEGAL HIST. 95, 96 (1986).
11. J.B. Ruhl, *The Fitness of Law: Using Complexity Theory to Describe the Evolution of Law and Society and its Practical Meaning for Democracy,* 49 VAND. L. REV. 1407, 1425 (1996).
12. Thomas C. Grey, *Langdell's Orthodoxy*, 45 U. PITT. L. REV. 1, 11 (1983).
13. Robert C. Berring, *Collapse of the Structure of the Legal Research Universe: The Imperative of Digital Information,* 69 WASH. L. REV. 9, 22 (1994).
14. *Id*. at 23.

painlessly."[15] With the digest system, "moderate, incremental reform remains quite possible but . . . foundational, transformative innovation [is] difficult."[16] Digests kept memetic variation to a minimum, encouraging the continued use of traditional legal principles and concepts while discouraging innovation. Memetic variation will be greater as CALR gradually displaces digests.

¶11 The traditional model or pattern of the law was established in the late 1800s with the digests. We have a vested interest in the digest organizational scheme. The digest scheme is like a grammar. Attorneys have been stuck on the digest organization as the perfect organizational pattern for the law, unable to see an alternate organizational pattern. "The categories contained in [the digest system] are like eyeglasses we have worn a long time. They enable us to see better, but lull us into thinking our vision is perfect and that there may not be a still better pair."[17]

¶12 According to Kuhn, a paradigm is a set of concepts, patterns, or assumptions to which those in a particular professional community are committed and which forms the basis of further research.[18] The textbooks studied by students training in a particular learned community contain the community's paradigm.[19] A paradigm shift occurs when some theories underlying the community's paradigm are discarded and are replaced with new ones.[20] This is usually preceded or accompanied by a recognition that the existing paradigm does not adequately provide answers to the problems considered by the community.[21] With the new paradigm, the community looks at the world from a different viewpoint and uses new instruments to solve problems.[22] At the time of a paradigm shift, some long-time members of the community will not welcome the transition.[23] Others, however, will accept the new paradigm because it is effective at solving problems or because it has greater elegance or simplicity.[24]

¶13 The result of a paradigm shift is that members of the community are no longer constrained by the framework of the old paradigm. The new paradigm allows them to use a new framework to organize data in an entirely different way. With the reorganization, members of the community may create startling new concepts or produce radically new solutions to problems.

---

15. Delgado & Stefancic, *supra* note 5, at 208 (citations omitted).
16. *Id*. at 217 (citations omitted). In the article, the authors focus on the Library of Congress subject classification system, the *Index to Legal Periodicals*, and the West digest system, referring to them as a triple helix and "the situation confronting the lawyer or scholar trying to break free from their constraints [as] the triple helix dilemma." *Id*.
17. *Id*. at 209.
18. KUHN, *supra* note 4, at 10–11.
19. *Id*. at 43.
20. *Id*. at 66.
21. *Id*. at 69.
22. *Id*. at 111.
23. *Id*. at 151.
24. *Id*. at 155.

¶14 In Kuhn's terms, the social sciences are prescience, lacking consensus on a central paradigm. The universal adoption of Langdell's conception of the law, however, provided a legal paradigm. Several authors now have already recognized CALR as the foundation of a new paradigm.[25] But before exploring the paradigm of the twenty-first century, the following section examines the characteristics that established the digests as the dominant legal research paradigm of the twentieth century.

Digests

Conservative and Orthodox

¶15 Robert Berring has characterized the West key number system as conservative and orthodox.[26] The conservative nature of the key number system is plainly shown by how slowly it has evolved over the years. Revisions in the digest scheme to reflect changes in legal terminology or in the law have occurred only after a considerable passage of time. For example, with the Ninth Decennial Digest (1976–1981), West renamed the topic "bastards" as "illegitimate children" and added the topics "abortion and birth control," "accountants," "bankruptcy," "condominium," "copyrights and intellectual property," "internal revenue," and "zoning and planning."[27] Later, West renamed "illegitimate children" as "children out-of-wedlock." West only added "drugs and narcotics" in 1972.[28] Although the topic "master and servant" could long ago have been renamed "employment law," the topic retains the archaic-sounding name even today.

¶16 The key number system is orthodox for two reasons. The first is that West staff attorneys are trained to edit a case in terms of the key number system; this applies even to a case given an unusual treatment by the judge authoring the opinion.[29] The second is that the West editorial process ensures that similar headnotes are consistently placed under the same key number. In that process, a staff attorney reads and prepares headnotes for each case that arrives at West. The attorney assigns each headnote to at least one topic and key number. At that stage, the case goes to senior editors who verify the placement of the headnotes within the West key number system, modifying the placement as needed. In the mid-1980s, all

25. Berring, *supra* note 3, at 29, 38 (calling the West reporter system the "old paradigm" and CALR the "new paradigm"); Barkan, *supra* note 9, at 636 (predicting that CALR will be a new paradigm).
26. Berring, *supra* note 3, at 36.
27. *Id.* at 35 n.29.
28. Bintliff, *supra* note 2, at 343.
29. Berring comments: "It is an interesting question whether West editors engaged in a kind of common law decisionmaking, classifying a case by inferring the 'proper' holding from the pattern of facts and the outcome, while downplaying the actual language of the opinion." Berring, *supra* note 3, at 33 n.26.

cases were funneled through four senior editors, who reviewed the placement of all headnotes.[30]

Use of the Digests

¶17 Digests present many advantages to the legal researcher. An expert editorial staff assembles the digests, spending significant time and effort organizing the information. Although this effort goes unnoticed by the average researcher, the extensive editorial work adds immensely to the value of the product. West publications are known for being accurate and reliable. The large number of digest topics makes the digest comprehensive in scope. All cases included in West digests, whether state or federal and whether trial or appellate, are indexed using the same key number system, which is used throughout almost all West publications. West publishes both the cases and the digests; its senior editors maintain the consistency of headnote categorization within the key number system. Given all this, digests are an essential case-finding tool.

¶18 Although digests have a number of advantages, there are almost an equal number of disadvantages to their use. The digest summary may be inaccurate, may take on a different meaning in the context of the entire case, or may cover dictum rather than the holding of the case. Classification of a summary under a particular digest topic and number is an unofficial decision made by the publisher; the text of the opinion supersedes any misleading digest classification. At the same time, an important point of law from a case may have been overlooked by the publisher and lack a corresponding headnote; details important to the researcher may be overlooked. And, as noted earlier, the key number system is extremely slow to react to changes in the law.

¶19 Using the same key number system for all jurisdictions, while advantageous in some ways, also presents problems for the researcher. Federal procedural and substantive law differs greatly from the procedural and substantive law of the individual states, and the law differs greatly from state to state. The complexity of the differences among the jurisdictions is masked by a universal key number system.

¶20 Finally, using the key number system can be time-consuming. The system is a deeply layered index; a researcher looking for a case first must locate the correct topic and then follow through all the layers in the outline before locating the case on point. On a more prosaic level, the need for a researcher gathering

30. *Id*. at 32. In another article, Berring lightheartedly likened the West editors to "a band of warrior monks, totally devoted to their mission. They spared no effort to produce a true and perfect product, they cared deeply about how people perceived the quality of their product. They saw no humor in what they did and they could be tough customers if someone threatened the holy doctrine. Like religious bands everywhere they were intolerant of criticism and would gladly pummel heretics, let alone impostors. But within their own code they were very true. It was a blessed consistency." Bob Berring, *Ring Dang Doo*, 1 GREEN BAG 2d 3, 4 (1997).

information to first use the digest index, then its volumes of case summaries, pocket parts, and supplements is a cumbersome process, especially when the researcher still must consult bound reporter volumes and advance sheets after the digest is used.

Indexing

¶21 To understand the difference between print research and CALR, it is necessary to understand print indexing and CALR indexing.

Print Indexing

¶22 An index allows the researcher to locate relevant information in a collection of documents. The most common type lists subjects alphabetically, followed by a reference allowing the researcher to locate the information in the document collection. Subject indexes tend to be document-oriented rather than request-oriented because of the process used to generate the index.[31] The individual generating the index is apt to use the wording of the document for the subject entries in the index rather than a term the researcher has in mind. Print indexes basically are oriented to the content of the documents, although they may also include a few multiple entries that attempt to anticipate the researcher's request.

¶23 The authority control list, called a thesaurus, is the master list containing all the possible subject entries in the index.[32] Terms in the index not included in the thesaurus are accompanied by an internal reference, cross-indexing the researcher to another subject entry from the thesaurus. Authority control performs two functions. First, it standardizes all the subject entries; all entries are fit under one of the subject entries on the authority control list. For example, the thesaurus for West digests normalizes all information concerning employment by placing it under the topic "master and servant" rather than under "employment." The researcher with an employment law question looking for the subject heading "employment" in the digest's descriptive word index is referred to "master and servant."

¶24 Second, authority control limits the number of subject entries in order to keep the index to a manageable size. The subject entry "employment," directing the researcher to "master and servant," occupies a single line in the index. The alternative is to expand the digest thesaurus to include multiple synonymous terms. For example, if the thesaurus contained both "employment" and "master and servant," identical information would follow each entry. Expansion of the digest thesaurus to include all possible terms a researcher might use in a search

31. *See* Daniel P. Dabney, *The Curse of Thamus: An Analysis of Full-Text Legal Document Retrieval,* 78 LAW LIBR. J. 5, 10 (1986).
32. *Id*. at 11.

query would significantly lengthen the digest and require repetition of much information.

¶25 Print indexes contain a linear, alphabetical listing of subject entries from the thesaurus. A subject entry may be followed by a number of subtopics; those subtopics may be followed by a number of sub-subtopics, and so on. The entries are printed in outline fashion, with the thesaurus subject entries printed at the left-hand margin and each successive level indented, with as many as four or five levels of outline stacked one on top of another.[33]

¶26 Digests are subject indexes that index legal principles rather than whole cases, subjects, pages, paragraphs, or words. West key numbers may correspond to the third or fourth level of the outline, and West key number subsections may correspond to the fourth or fifth level of the outline. The multivolume descriptive word index of a digest adds an additional layer or more to the stack. The heavy stacking allows each of 20 million headnotes from an estimated 5 million cases to be assigned to one of over 100,000 key numbers.[34]

CALR Indexing

¶27 The most obvious difference between print indexing and CALR indexing is that the latter is constructed by the researcher rather than by the publisher. The success of a CALR search depends on recall, precision, and fallout. The *recall* percentage is the number of relevant documents retrieved compared to the total number of relevant documents in the database. The *precision* percentage is the number of relevant documents retrieved compared to the total number of documents retrieved. The *fallout* percentage is the number of irrelevant documents retrieved compared to the total number of irrelevant documents in the database.[35]

¶28 To increase recall, the researcher broadens the scope of the search. With a broader search, recall increases but precision decreases, since such a search retrieves more documents, many of them irrelevant. Many CALR databases contain a large number of documents and the fallout problem can be significant. To increase precision, the researcher might limit the number of documents retrieved by adding keywords to the search query. Of course, with a narrower search, precision increases but recall decreases, possibly omitting a number of highly relevant documents.

¶29 Precision is an easy percentage to measure but recall is not, since the researcher would have to know the total number of relevant documents in the database and compare the number of relevant documents found using a particular search query against that total. A curious phenomenon is that researcher percep-

33. *Id*. at 12–13.
34. *West Group Reclassifies Insurance Key Numbers*, NAT'L L.J., Mar. 20, 2000, at B11 (advertisement); Bintliff, *supra* note 2, at 343.
35. *See* Dabney, *supra* note 31, at 15–17 (discussing the concepts of recall, precision, and fallout).

tion of the thoroughness of CALR varies greatly from the recall actually achieved. A researcher, happy with retrieving a number of relevant documents, may overestimate recall and be falsely confident that the research results are satisfactory. "A relatively high precision will give high user satisfaction as long as alternative methods do not disclose to the lawyer that the recall is low. The lawyer is satisfied and the providers of the system get positive feedback from their users."[36] Several studies have documented this "false confidence syndrome."[37] An often-referenced study indicates that the average recall may be as low as 20%.[38]

Computer-Assisted Legal Research

¶30 CALR, once limited to information available online from Westlaw and LEXIS, now includes information from other commercial publishers and the Internet. To use it successfully, the legal researcher must understand CALR generally and the specific characteristics of Westlaw, LEXIS, and Internet search engines.

Use of CALR

¶31 Perhaps the most important thing for a researcher using CALR to understand is that a relevant document must exactly match the search query or it will not be retrieved. A CALR search is literal. If the query contains a misspelled word, the computer will locate any instances of the word misspelled in the same way, but it will not find instances where the word is correctly spelled (presumably the spelling that would occur in the greatest number of relevant documents). The search query also may locate unanticipated, ambiguous uses of search words. Often, the true meaning of a passage is implicit rather than explicit or the author uses metaphors or figurative language. A metaphorical, figurative, or implicit reference to a key legal concept, especially if the concept is abstract, is virtually impossible to locate using full-text retrieval.

¶32 Formulation of the query can be difficult if the keywords have many synonyms, can be stated in many different ways, or can express several different ideas. Statutes are particularly difficult to research because their language is often both repetitive and unique. At the same time, the target statute may contain a distinctive, but unanticipated term. Researching procedural questions, abstract concepts, and legal rules is difficult, since many of the keywords tend to be ones that

---

36. Jon Bing, *Performance of Legal Text Retrieval Systems: The Curse of Boole,* 79 LAW LIBR. J. 187, 197 (1987).
37. *See,* e.g., F.W. Lancaster et al., *Searching Databases on CD-ROM: Comparison of the Results of End-User Searching with Results from Two Modes of Searching by Skilled Intermediaries,* 33 RQ 370 (1994).
38. The study and its results, David Blair & M.E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System*, 28 COMM. ACM 289 (1985), are described in Berring, *supra* note 3, at 43–46, and in Dabney, *supra* note 31, at 28–29.

commonly occur in legal documents and the relationship between individual common words is complex. A novel research question presents a great challenge if the researcher is searching for cases with facts or concepts that can be analogized to the facts or concepts in the question.

Characteristics of LEXIS and Westlaw

¶33 Westlaw and LEXIS contain the full text of a massive number of documents. The researcher uses Boolean logic to retrieve relevant documents.[39] Boolean logic is a "syntactical calculus," a mathematical algebra that allows the researcher to tailor the search query by using the Boolean operators "and," "or," and "not." The researcher also can specify the proximity of keywords within the search query, requiring that one search word be within the same sentence or paragraph (grammatical connectors), or within a specified number of words (a numerical connector) of another search word. In addition, the researcher can restrict the search query to particular recurring elements of a case or statute, such as the citation of a case, the names of the parties in a case, the court issuing the opinion, the judge authoring the opinion, the date of the opinion, the attorneys participating in the case, the case syllabus, the citation of the statute, the text of the statute, statute headings, historical references, and research references. In Westlaw, the researcher can limit the search query to the headnotes and the digest topics.[40]

¶34 Westlaw and LEXIS documents are located through a concordance. The concordance is an index invisible to the researcher containing an alphabetical listing of every word and every root word in the documents, except for extremely common words like "an," "of," and "the." Each word in the concordance is accompanied by its location within the database, by document, paragraph, sentence, and position within the sentence. The concordance also identifies the data corresponding to a recurring element of a case or statute, such as a citation or party name. When the researcher submits a search request, the computer matches the keywords or phrases to the words in the concordance. This shortens the search time, allowing the computer to search the alphabetical listing in the concordance rather than the full text of each of the documents in the database.[41]

Natural Language

¶35 Westlaw and LEXIS both offer a "natural language" alternative to Boolean search queries. This approach allows the researcher to submit a query without using Boolean connectors and without having to anticipate the exact words used

39. The term Boolean logic was named after the British mathematician George Boole (1815–1864). CHRISTOPHER G. WREN & JILL ROBINSON WREN, USING COMPUTERS IN LEGAL RESEARCH: A GUIDE TO LEXIS AND WESTLAW 23 n.20 (1994); Bing, *supra* note 36, at 190.
40. *See* WREN & WREN, *supra* note 39, at 114–24.
41. Berring, *supra* note 3, at 41 n.44; Dabney, *supra* note 31, at 17 nn.19–20.

in the relevant document. Relevancy is based on two assumptions. The first is that a word that appears relatively infrequently in the database has the same meaning each time it appears. Therefore, if that word is part of the search query, the presence of the word in a document is significant. The second assumption is that the more times a search term appears in a document, the more relevant is the document.[42]

¶36 The natural language program processes the search query in a series of steps. It first identifies the keywords in the search query. Then it eliminates frequently used words from the query. In the third step, the program strips keywords to their roots and generates derivatives of those roots. Next it retrieves documents that match concepts in the search query, performs a complex statistical analysis on the documents retrieved, and ranks the retrieved documents for relevancy.[43] The researcher has the option of using the computer thesaurus to add alternative terms to the search query.[44]

¶37 Although natural language searching may be attractive because of its apparent ease of use, one study finds that it is not markedly better than Boolean searching and recommends using both types of searches to obtain the best results.[45]

Characteristics of Internet Search Engines

¶38 Internet search engines employ automated programs to gather information from Web pages and then use that information to construct an index. The automated programs are referred to as "webcrawlers," "crawlers," "spiders," "robots," and "bots." The webcrawler revisits a site on a regular basis, such as every month or two, to find new information. Depending on the search engine, the crawler may collect information from the full text or from certain portions of a Web page or Web site. The information it collects includes the content of the Web pages visited and information contained in meta tags[46] provided by the creater of the Web site. The information collected is compiled into an index.

¶39 When a researcher submits a search query, the search engine matches the

---

42. WREN & WREN, *supra* note 39, at 26.
43. Sheilla E. Désert, *Westlaw Is Natural v. Boolean Searching: A Performance Study*, 85 LAW LIBR. J. 713, 715–16 (1993).
44. *Id*. at 720.
45. *Id*. at 741–42. Two other authors believe that natural language searching will not displace Boolean searching. WREN & WREN, *supra* note 39, at 26. One author compared Boolean searching to manual transmission on a car and natural language searching to automatic transmission. Both types of cars transport passengers to a destination, but the manual transmission allows the expert, the race car driver, to more finely tailor the car's performance on the drive. Posting of Russ Armstrong to lawlib@ucdavis.edu (Jan. 8, 1996), *quoted in* Bintliff, *supra* note 2, at 347.
46. A tag is a line of HTML code invisible to the Web page viewer. A meta tag is a tag containing keywords. The meta tag information is invisible to the researcher viewing the Web site unless the researcher reads the HTML code. A Web site usually contains tags for the title of the site, the URL, a description of the site, and keyword tags.

terms in the query to terms in the index and ranks the Web sites found for relevance. Search engines use a complex statistical analysis, much like the search engine in natural language, to determine relevancy based on the location and frequency of keywords. The search engine will consider a Web page containing the keywords in the title or near the top of the page or containing frequent use of the keywords relevant to the search.[47]

¶40 Web search engines are generally less sophisticated than Westlaw and LEXIS search engines. The capability of search engines varies widely and the use of Boolean operators is not standard from one search engine to another. Although many Web search engines allow use of Boolean connectors, one expert advises against their use, calling them "overkill for the average web user."[48] He offers this advice because search engines present their results with the most relevant first; the highest ranked documents are usually the ones in which the search terms most frequently occur.

¶41 For example, the only Boolean connectors available with lawcrawler (http://lawcrawler.findlaw.com) are "and," "or," "near," and "not." The connector "near" allows a search for two words within ten words of each other. Thus, proximity searches are much more limited than they are in Westlaw and LEXIS. The researcher cannot require that one search word be within the same sentence or paragraph, or within a specified number of words of another search word (other than within ten words). As on Westlaw and LEXIS, the researcher is generally more successful using uncommon, unambiguous, and infrequently used words.

A Comparison of Print and Full-text CALR Research

¶42 Legal research requires a combination of factual knowledge and higher-level intelligence. Factual knowledge involves recognition of objects, activities, and locations. A higher level of intelligence enables the researcher to create, manipulate, and apply abstract concepts to the given facts. High-level perception involves recognition of relationships and abstract ideas and concepts. It draws meaning out of objects, activities, and locations. This type of insight allows the researcher to think in the abstract, recognizing patterns in the facts, issues, and primary sources and regrouping them to recognize new patterns. The researcher may recognize similarities, distinctions, and relationships. Out of these patterns, insight may allow the researcher to notice analogies and build abstract conceptual legal frameworks.

---

47. For a recent article on using search engines for legal research, see Deanna Barmakian, *Better Search Engines for Law*, 92 LAW. LIBR. J. 399, 2000 LAW LIBR. J. 36.

48. Danny Sullivan, *Search Engine Watch: Boolean Searching, at* http://searchenginewatch.internet.com/facts/boolean.html (last visisted June 2, 1999), *quoted in* Jessica R. Hogan, *"Why Won't My Westlaw Search Work on Lycos?"* 7 PERSPECTIVES: TEACHING LEGAL RES. & WRITING 123, 125 (1999).

¶43 Imagine a researcher presented with a legal problem in an area of the law with which the researcher is unfamiliar. First, the researcher will analyze the facts. Then, if using print sources, the researcher may consult secondary sources such as legal encyclopedias or law review articles to learn about the legal concepts basic to that area of law. After consulting secondary sources, the researcher might use keywords to locate relevant primary sources. After locating these, the researcher will synthesize the primary sources, extract a legal principle, and determine how the legal principle applies to the legal problem.

¶44 To summarize, the legal researcher using print sources:

1. analyzes facts in the legal research problem;
2. evaluates what legal concepts may be relevant;
3. researches concepts in secondary sources;
4. finds primary sources concerning relevant concepts;
5. synthesizes the principle contained in the primary sources; and
6. applies the principle to the legal research problem.

¶45 How is this different from the steps taken by the legal researcher using CALR? The CALR researcher identifies key terms in the legal problem and develops a research query. Using the query, the researcher locates cases with similar facts and other applicable primary sources. The researcher synthesizes the primary sources and extracts a legal principle. At this point, some researchers may analyze how the principle fits into the overall framework of that area of law. Other researchers may stay at the factual level, never taking the time to analyze how the principle fits into a larger overall scheme. Then the researcher will apply the principle to the legal problem.

¶46 Thus, the legal researcher using computer-assisted legal research:

1. analyzes facts in the legal research problem;
2. identifies keywords;
3. locates applicable primary sources;
4. synthesizes the principle contained in the primary sources;
5. may analyze how the cases fit (or do not fit) into relevant legal theories and public policy arguments; and
6. applies the principle to the legal research problem.

¶47 An attorney researching in print sources typically travels from the facts of the research problem to general concepts and then to the specifics. Secondary sources may be used to learn about an unfamiliar area of law. While using digests, the researcher learns about the traditional organization of the area of law and how a relevant case fits into that pattern.

¶48 An attorney researching online typically does a word search, looking for cases containing the same facts. If the search retrieves a number of cases with similar facts, the attorney may be satisfied with the outcome. However, a search that discovers factually similar cases does not also offer a theory of law as its nat-

ural result. Additional work and creative energy on the part of the researcher are required to formulate a legal theory. The CALR researcher who remains focused on facts may neglect broader issues and legal concepts, and may be oblivious to the general perspective. When using digests, however, the location of the relevant case summary within the digest system has meaning in and of itself and helps the researcher see the context of the case. With CALR there is no immediate context for a relevant case. The attorney may not consider the importance of the context of the facts or the role that fairness or justice might play in making a persuasive case for the client. The attorney may remain at the factual level, failing to consider legal concepts or public policy arguments.

¶49 The computer format itself may exacerbate the CALR researcher's tendency to remain at a factual-level analysis of retrieved cases. The computer generally only allows a small amount of information to be accessible on each screen, whereas print sources make a much greater amount of information instantly available, allowing the reader to glance over two pages of information at a time and quickly leaf through a case to determine its relevancy. The researcher may read cases online rather than printing them out and studying them in detail. The researcher may even use the copy feature to copy portions of a case and paste them into a legal document without thoroughly examining their context or determining how they fit into broader legal concepts.

¶50 The expert legal researcher is creative and uses good legal analysis. Whether researching in print sources or using CALR, the most successful researchers go beyond established legal theories to create new legal concepts. CALR presents a new challenge to the legal researcher. CALR will differentiate between the mediocre researcher and the expert researcher. The expert researcher is willing to think outside the box.

The Paradigm Shift to the Code

¶51 We are ready for a paradigm shift. The first versions of CALR software signaled a shift in responsibility from the commercial publisher to the researcher and shortened the distance between the researcher and the primary sources. However, CALR provides a software interface between the researcher and the database. The interface is invisible to the researcher, yet the more sophisticated the software, the greater the distance between the researcher and the data.

¶52 Berring notes what he calls "the transparency of research protocols."[49] The average attorney who uses CALR does not have the patience or time to become thoroughly trained or read all the instructions. The attorney will more readily use a CALR system that can be easily navigated with a minimum of training.

---

49. Robert Berring, *Chaos, Cyberspace and Tradition: Legal Information Transmogrified,* 12 BERKELEY TECH. L.J. 189, 209 (1997).

"The bulk of research and development in this field centers on developing easier and quicker means of using the systems."[50] Berring relates a candid revelation made to him by a CALR vendor representative, to the effect that "[t]o invest $5,000,000 in implementing a new search system could demonstrably improve research effectiveness, but that might raise costs and might introduce a new layer of difficulty for the user and thus would be counterproductive. The money would be better spent in marketing."[51]

¶53 Berring points out that the more automatic a search system becomes, the less the researcher knows about how documents are retrieved, and the less the researcher is capable of directing the process.[52] An illustration of this difference is a keyword search using Boolean logic in Westlaw or LEXIS versus a natural language search. As previously described, the researcher using Boolean logic has more control over the retrieval process. A natural language search query places the researcher at a greater distance from the retrieval process. Rod Borlase counsels against using natural language queries because they add an additional layer of software between the researcher and the database, achieving "a poorer result and greater expense."[53] He points out that even a nonsense search query like "Do lasers eat butterflies or just seduce bridge builders?" will retrieve twenty documents.[54]

¶54 The digest system represents the present paradigm of legal research. Gradually, we will be shifting to a new legal research paradigm as a greater and greater percentage of research is accomplished on the computer. The new legal research paradigm will be the CALR protocol to which Berring referred; it is often called "the code" by the technologically savvy.

¶55 Attorneys usually think of law as comprised of primary sources—constitutions, statutes, cases, and administrative regulations. Professor Lawrence Lessig of the Harvard Law School presents the message that "code is law." But Lessig is not speaking so much of legal codes as of computer codes.

¶56 In *Code and Other Laws of Cyberspace*,[55] Lessig predicts that "code" will assume a role like that of traditional primary sources in cyberspace.

> [T]here is regulation of behavior in cyberspace, but that regulation is imposed primarily through code. What distinguishes different parts of cyberspace are the differences in the regulations effected through code. In some places life is fairly free, in other places controlled, and the difference between them is simply a difference in the architectures of control—that is, a difference in code.[56]

---

50. *Id*.
51. *Id*.
52. *Id*. at 209–10.
53. Rod Borlase, *Learn Westlaw & Lexis/Nexis Essentials! (Or, What They Will Tell You <Maybe> Only If You Ask), at* http://www.law.uh.edu/guides/essentials.html (last visited Jan. 22, 2001).
54. *Id*.
55. LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE (1999).
56. *Id*. at 20.

Cyberspace code is not enacted into law by an elected legislature. Instead, it is crafted by code writers.[57]

¶57 For Lessig, code is the hardware and software that creates the architecture of cyberspace.[58] Code is "the 'built environment' of social life, its architecture."[59] Code is powerful. Lessig calls it the "newly powerful regulator in cyberspace."[60] Code defines the dimensions of cyberspace; it enables one traveling through cyberspace to take certain actions and constrains the traveler from taking other actions. "Architecture is a kind of law: it determines what people can and cannot do."[61]

¶58 Code has nothing to do with law. The search is for strings of characters and not for meaning. Code writing is becoming commercial. Most attorneys will be pleased with new legal research software that allows research without training time. They will welcome such software without questioning the reduced control they will have over their research. Most attorneys think that legal research software is just a question of computer engineering. Code transfers indexing from lawyers to programmers and administrators (managers at Westlaw and LEXIS or administrators of legal research sites on the Internet).

¶59 The software capabilities are left to the major online publishers. In the past, West had control of indexing through its comprehensive digest system. While control through use of the digests may be lessening, West Group and LEXIS Law Publishing will control the code in Westlaw and LEXIS software, and administrators of other online legal research sites will determine the code for those sites. The large legal publishers will remain powerful, but the power will shift from the print digests to research software. The code writers are employees of West and LEXIS, controlled by the employer. Materials will be indexed by nonlawyers who have the requisite code literacy. Indexing was apparent with digests but is not readily apparent with software.

¶60 West Group and LEXIS Law Publishing can change the behavior of legal researchers by changing the code architecture of their software. The code architecture is set by the code writers, under the direction of management. Choices will be based largely on profit. Which architecture will be the most appealing to legal researchers and produce the most profit? Choices may be dictated differently on a government or noncommercial site.

¶61 Code architecture constrains by channeling the research process and the behavior of the legal researcher. Code conditions the researcher's access to online data. It does this by facilitating certain actions and making others impossible.

57. *Id*. at 53.
58. *See id*. at 6.
59. *Id*. at 86 (citations omitted).
60. *Id*.
61. *Id*. at 59.

Legal research code assumes certain research behaviors and facilitates them. Other behavior might fall outside what the code writers expect the researcher to do. Behaviors outside the code writer's expectations may be difficult or impossible. Code channels are hidden, made invisible to the researcher by a systems graphic interface. Thus, code is a persistent and unrecognized influence that channels online research. It is like a maze in that the researcher may take certain routes, but not others which are sealed off by the maze walls, often invisible to the researcher. Code determines the possibilities of Westlaw and LEXIS; the researcher does not have the power to change the code provided by either system.

Conclusion

¶62 Most lawyers and judges were trained in the Langdellian case method, which attempted at all times to reconcile new situations with old principles, only departing from old principles when they clearly no longer served changed circumstances. Soon, this will be supplanted with CALR. Most lawyers and judges will learn to disdain the patriarchal elitism of Langdell's pedagogy; and, more important, they will come to rely on the Sphinx-like computer for their answers. They will submit their queries to an increasingly mysterious, but increasingly efficient, machine. Concepts like justice, fairness, and equity mean no more to this machine than "silicon breast implant." In fact, the query "silicon breast implant" will result in a far more coherent set of responses than "due process" because the machine performs best with very specific queries. We assume that, despite the nature of the responses, lawyers and judges will still argue principles and justice in resolving disputes. But we are definitely moving in another direction.

¶63 The danger of the present situation has been noted in poignant analysis by Lessig. He addresses three legal areas: intellectual property, privacy, and free speech. He is alarmed at the threat the code poses to each of these, but he does not press his values upon us. In fact he states his preferences rather reluctantly, because his purpose is primarily to make us aware that the code the computer uses readily subordinates the values we may attribute to these legal questions to the values embodied in the coding of the computers, especially the Internet.

¶64 Lessig's point is that the code that runs the computer has its own rules that inevitably become our law. The point is more obvious when we acknowledge that the rules by which computers communicate, although founded in the practice of engineering, can be regulated in a variety of ways that affect our lives deeply. If, for example, our abhorrence of child pornography is challenged by ready access to it on the Internet, we can fight child pornography in a number of ways. The most effective way might subject all of us and all our communications to electronic monitoring. The government might spy on us at any moment without our awareness. We could have regulation that would allow the government access to our computers at any time. The infrastructure of the Internet has a code, a code

that can be altered and a code that has its own laws, which are our laws as well. We are not accustomed to perceiving law in this way, but Lessig provides a compelling argument for that perception. The choices that are made in the code are likely to impose powerful limitations on privacy and freedom of speech. Regulation is accomplished through the code rather than by traditional legal means.

¶65 Frightening though Lessig's vision may be, the full story has yet to be told. The code controls even the means of knowing the law. As legal research depends more and more on computer searches, the law becomes more and more a product of the code and less and less a product of human judgment. The old system of indexing and the Langdellian pedagogy created an image of law drawn from original sources. Indexing by its very nature made law more orderly, more logical, more faithful to its origins. Computer indexing by word or character treats all laws as equal. Nothing inherent in the computer argues that any case or statute, principle or policy has hegemony over any other. The result is clearly postmodern—anyone's feeling about a dispute is on an equal basis with that of any other person. "Justice" has the same code status as "implant." After all, they have the same number of characters; they will be treated with the same reverence. Some might argue that this system puts uncensored data in the hands of the researcher, who makes up his or her mind as to the meaning and importance of the data that is retrieved. This is raw data, after all, without the benefit of any guiding hand. It is very democratic because it is chaotic.

¶66 Language approximates ideas. There is no exact correlation between language and ideas. An individual may use different words at different times to express the same idea. Different individuals use different words to express the same idea. Yet CALR treats legal materials as one vast sea of undifferentiated data, distinguishable by the character strings floating in that sea. A query will return a set of materials that surround those character sequences. CALR returns us to a more primitive legal regime in which "the law" consisted of a list of the laws. The primitive attitude toward literacy likened similar characters to similar meanings. Dog is a sacred animal because it is god spelled backward. With CALR, all data is equal and search queries retrieve matching character strings.

¶67 Paradigm shift may seem too strong a characterization of this change. After all, it was invented by Kuhn to describe the movement in a science to a revolutionary new consensus. But what could be more important in the law than the reordering of our hierarchy of legal values?