

FULL-TEXT DATABASES AND LEGAL RESEARCH: BACKING INTO THE FUTURE

BY ROBERT C. BERRING †

INTRODUCTION

The use of computers in legal research is a topic at the center of any discussion about modern legal literature. The appearance of computer terminals at each accredited law school and the presence of on-line systems at every large law firm signal a major change in the way that lawyers conduct research. As more and more print publishers introduce on-line versions of their traditional hard copy products,¹ the legal profession is increasingly receptive to a primary role for computers in legal research.² This article will examine one crucial role for the computer—the full-text on-line computer databases.³

© 1986 *High Technology Law Journal*

† Professor of Law and Law Librarian, Boalt Hall School of Law, University of California, Berkeley; Dean, School of Library and Information Studies, University of California, Berkeley. I wish to thank my colleagues Daniel Dabney, Bill Maron, and Patrick Wilson for helpful suggestions, and my research assistant Kathleen Vanden Heuvel for her invaluable help. Also, I wish to thank *High Technology Law Journal* editors Mark Ryland, Cecelia Fusich, and Jeff Love for their contributions to this article.

1. See, e.g., BNA LABORLINE, an on-line product by the Bureau of National Affairs similar to its looseleaf service. Commerce Clearing House and Prentice Hall are also preparing on-line products that cover the same territory as the companies' loose leaf products.

2. See, e.g., Yates, *Nearly Everything You Want to Know About Data Bases*, A.B.A.J., Nov. 1985, at 90 (includes a list of over 70 databases "for virtually any area of law or any area related to the practice that you can imagine").

3. There are two important and related topics that I will not consider. First, the use of non-legal databases in legal research is growing. For example, many lawyers are beginning to use NEXIS, a database of Mead Data Central, that includes all on-line information services available from Mead except for those provided to the legal community as LEXIS. The NEXIS database consists of the full text of 142 newspapers, magazines, wire services and newsletters. NEXIS is available through the LEXIS system. See MEAD DATA CENTRAL, INC., *GUIDE TO NEXIS AND RELATED SERVICES* (1985).

Second, the use of computers in law office management and litigation support is burgeoning. See *Legal Times*, Sept. 30, 1985 (Special Supp. *1985 Fall Law Office Equipment and Services Directory*), which lists an enormous variety of computer services. Whole periodicals such as *Computer Lawyer* and *PC Lawyer* are devoted to this topic, and even established periodicals such as the *American Bar Association Journal* have continuing columns on law office computing.

The giant on-line databases produced by Mead Data Central (LEXIS) and the West Publishing Company (WESTLAW) are the main arena for the face-off between lawyers and computers. They have grown from simple repositories of case law into integrated databases of primary and secondary source materials.⁴ The systems are full-text⁵ and allow free-text searching⁶ for any word or combination of words, with query structures that incorporate the power of Boolean logic.⁷ This combination of sources and search capabilities creates an entirely new genre of legal literature.

LEXIS and WESTLAW already have become an integral part of the arsenal of research tools available to the lawyer, but we need to re-evaluate the role that they play in legal research. Recent studies have raised questions about both the general efficacy of full-text systems, and the research skills necessary to use them efficiently. These questions must be examined in order to assess the usefulness of the legal databases. One purpose of this article is to explore some problems in compu-

4. The legal databases now include various citation services, numerous legal periodicals, libraries for specialty practices, ALR annotations, etc. In this article I will focus on the use of the databases to search for judicial opinions, but much of what I will say could be applied to other on-line source materials.

5. A "full-text" database is one that incorporates every word of every document rather than the more usual (and less expensive) method of putting only an index entry or abstract of the document on-line.

6. "Free-text searching" enables the researcher to search for every occurrence in the database of any word or combination of words without using a pre-existing index.

7. "Boolean logic" is a syntactical calculus used for the comparison of data items (words or numbers) and combinations of data items. In Boolean logic, data items can be related in only one of two ways: true (matched) or false (not matched). For purposes of searching, one data item can be combined with others using the Boolean operators "and," "or," and "not." With the use of conjunctions, disjunctions, and negations, a search can list instances in a database where a given item or a combination of items exists.

The power of a Boolean search is this ability to match items that have specific relationship within a document. In a full-text search system, such as LEXIS or WESTLAW, the use of these conjunctions allows the researcher to create a context — to specify a relationship between the terms for which he is searching. For example, without conjunctions, the searcher would use one term at a time, calling up every instance of the word in a database. He would then have to examine each of these retrievals in order to discard those items not relevant to his issue. The use of a conjunction, "and," would allow him to search for instances where two words (or numbers, etc.) are found in a single document, paragraph, or sentence. The conjunction allows the researcher to specify a relationship between two terms, and thus formulate a more precise search.

The advantage of this search technique over a prepared index is that the researcher can find every occurrence of a significant word. This allows the researcher to both narrow the search to specifics and to broaden it, as he or she does not rely on the preselection of certain cases or sections of a document by the individuals who created the prepared index. This method also has inherent weaknesses. See *infra* text accompanying notes 48-86.

ter-based research, and to point out certain significant limitations of on-line full-text legal databases.

But we cannot examine the utility of legal databases in isolation. The impact of LEXIS and WESTLAW is not simply a matter of a new technology simplifying or speeding up a preexisting process; it involves a change in the structure of legal literature. More work needs to be done on the relation between the structure of legal literature and the substantive development of law,⁸ but it seems clear that in law, more than any other discipline, the structure of the literature implies the structure of the enterprise itself.⁹ I will attempt to show this interrelation by first describing the history of the traditional hard-copy primary sources, and assessing their influence on the lawyering process. Next, I will examine the emergence and growth of the legal databases. Finally, I will point out certain practical problems inherent in full-text searching and in making each lawyer his or her own on-line researcher, and suggest some theoretical difficulties with this new form of legal literature.

I. THE STRUCTURE OF THE OLD PARADIGM: THE WEST REPORTER SYSTEM

Before the arrival of computerized legal research in the 1970s, American legal publishing was a highly integrated and well-developed system of comprehensive publication and retrieval in hard copy. Most aspects of this system can be attributed to a few enterprising publishers who conceived of the intriguing publication formats. A brief sketch of the hard copy system is necessary to put the advent of the legal databases in context.

A. The Development of the System: Comprehensive Regional Reporters

The publication of case reports was organized, systematized and perfected by the West Publishing Company of St. Paul, Minnesota at the end of the nineteenth century.¹⁰ John B. West, an entrepreneurial office

8. Cf. Childress, *The Hazards of Computer-Assisted Research to the Legal Profession*, 55 OKLA. B.J. 1531 (1984) (suggesting certain tangible and intangible links between the structure of legal literature and styles of practice).

9. Recall Langdell's famous aphorism, "The library is to us what the laboratory is to the chemist or the physicist and what the museum is to the naturalist." HARVARD LAW SCHOOL ASS'N, *THE CENTENNIAL HISTORY OF THE HARVARD LAW SCHOOL 1817-1917*, at 97 (1918).

10. See W. MARVIN, *WEST PUBLISHING COMPANY: ORIGIN, GROWTH, LEADERSHIP* (1969) for an extensive if over-flattering portrait of the history of the West Publishing Company. For a more balanced treatment, see Woxland, "Forever Associated with the Practice of Law": *The Early Years of the West Publishing Company*, LEGAL REFER. SERV. Q., Spring 1985, at 115-24.

supply salesman, noted the disorganization of the case reporters that his lawyer-customers were purchasing. The existing forms of publication were slow, unorganized, and inaccurate.¹¹ In response he began to publish *The Syllabi*, a reporter that contained the text of Minnesota Supreme Court cases and summaries of decisions from surrounding states. *The Syllabi* was so successful that West introduced a successor entitled *Northwestern Reporter*.¹² This series contained the full-text of all decisions from those states that Mr. West considered the "northwestern" region. Publication was frequent, and the reporter was inexpensive and reliable. It was a success.

West soon realized that a "regional" reporter, which gathered together the decisions of a variety of jurisdictions into a single series, was useful to lawyers and, consequently, easy to sell. One of the motivations for the regional reporter approach was to gather enough cases to produce a regional biweekly advance sheet that was marketable to lawyers in a number of different jurisdictions. These advance sheets, which might have been too costly to produce for limited circulation in each individual jurisdiction, delivered judicial opinions quickly into the hands of the lawyers throughout the region. West made agreements with various courts to obtain decisions directly and rapidly, and spared no effort to locate opinions. His deserved reputation for completeness as well as accuracy rapidly earned him a substantial following.

Northwestern Reporter was only the first step in West's process of innovation. Mr. West extended his system nationwide by dividing the entire country into seven regions and by producing a reporter for each region.¹³ Within a few years, West Publishing Company provided comprehensive coverage of all state cases. The introduction of *Federal Reporter* and *Supreme Court Reporter* in 1886 completed this pattern. Although more established publishers also created regional reporters,¹⁴ no one emulated West's decision to divide the entire nation into seven

11. Young, *A Look at American Law Reporting in the 19th Century*, 68 LAW LIBR. J. 294 (1975), provides an overview.

12. The title *Northwestern Reporter* was actually used to describe two separate publications. The first *Northwestern Reporter*, introduced in 1877, was an enlargement of *The Syllabi*. It was still more like a newspaper and included the full text of only Wisconsin and Minnesota decisions. The *Northwestern Reporter* as we know it first appeared on April 26, 1879. F. HICKS, MATERIALS AND METHODS OF LEGAL RESEARCH 145-46 (3d ed. 1942).

13. John B. West's division of the country into seven geographic regions demonstrated his talent as a publisher but it did not show him to be prescient as a geographer. Ironically, West did not anticipate the development of the West; Oklahoma is not considered by many a Pacific state.

14. For example, *New England Reporter*, *Central Reporter* and *Western Reporter* by the Lawyer's Co-operative Publishing Company of Rochester, New York. See W. MARVIN, *supra* note 10, at 48.

regions and to provide coverage of all state court decisions. Indeed, contemporary commentators lambasted the idea as being wasteful and greedy.¹⁵

But it was this national coverage that most historians regard as the foundation of West's success.¹⁶ Lawyers were entranced by the availability of cases from all jurisdictions in a standard, inexpensive format. Whether by dint of the product's attractiveness, its price, or its marketing, the National Reporter System was a resounding success. Other competitors soon dropped out of sight and left West with a dominant position as the unofficial publisher of cases.¹⁷

Although official reporters sponsored by the various jurisdictions continued to exist,¹⁸ the comprehensive West publication system became the most prominent feature of American case reporting. During the same period the number of cases being rendered into written opinions was rapidly increasing,¹⁹ lending impetus to West's scheme. Moreover, West's traditionally high standards of speed and accuracy in publishing enhanced his system's reputation and marketability. The fact that the West regional reporter structure has thrived until today is testimony to its quality and usefulness.

B. The Structure of the System: Headnotes and the American Digest

The American Digest System²⁰ was the key aspect of the new form of legal literature that Mr. West created. The Digest classified all areas of law into seven broad categories. These categories were then subdivided into some four hundred and thirty topics. Each topic was then further subdivided into subsections called "Key Numbers" (a trademarked term). These Key Numbers allowed the topic to be broken into as many subdivisions as were necessary to completely cover that area of

15. *The New Reporters*, 19 AM. L. REV. 932 (1885).

16. See, e.g., Woxland, *supra* note 10, at 116.

17. See *id.* at 122.

18. Today, for example, only 29 states publish official reporters at all, and many only include cases from the highest court of that jurisdiction. See HARVARD LAW REVIEW ASS'N, A UNIFORM SYSTEM OF CITATION 136-76 (13th ed. 1981). The only officially published federal cases are the Supreme Court cases in *United States Reports*.

19. The literature bemoaning the volume of published cases is vast. A personal favorite is High, *What Shall Be Done with the Reports*, 16 AM. L. REV. 435 (1882).

20. In 1889, West acquired *U.S. Digest* from the Little Brown Company and editor Benjamin Vaughn Abbott. *U.S. Digest* was modified and published by West as its American Digest System. Even more important was the acquisition of *Complete Digest* because its editor, John Mallory, also came to West. Mr. Mallory is acknowledged in the preface of *First Decennial Digest* as the guiding hand behind the American Digest subject scheme.

the law.²¹ Eventually, a structure of subject headings was created which provided for *every* possible legal issue. A headnote always had a specific location in the Digest System.²²

West Publishing Company developed an elaborate process for melding the cases into the Digest. As cases arrived, a lawyer-editor read each one, editing it first for citation form and other stylistic conventions. Then the editor prepared a set of headnotes that served as abstracts of each point of law contained in the decision. Each headnote was assigned to a specific topic and Key Number location.²³ A headnote could be assigned to two locations but it *had* to fit into at least one Key Number address. The text was then passed to one of West's four senior editors who verified the accuracy of the topic and Key Number assignments. Although no statistics are kept, according to my conversations with senior editors, a substantial number of topic and Key Number assignments were modified at this point. The importance of the placement of the headnote into the Digest's subject index cannot be overemphasized. This initial placement had a tremendous impact on any subsequent manipulations of the data. In recent advertising, West indicated its internal valuation of the senior editors by calling them "Edi Knights."

The West Digest System exemplified a type of index called a universal subject thesaurus.²⁴ The concept of a universal subject thesaurus, while not unusual in information science, reshaped legal research. For when West Publishing created the Key Number System, it not only enabled lawyers to research cases by subject, it also allowed and encouraged lawyers to fit every legal issue into a certain conceptual framework. At a mechanical level, the West Key Number System created a comprehensive subject format that allowed for all of the cases appearing in the National Reporter System to be arranged by subject according to their headnotes; the power of the system made it the primary

21. For a classic description of West's American Digest System and a list of the categories and topics, see F. HICKS, *supra* note 12, at 233-43.

22. The practical and theoretical implications of this closed-end system are explored below. See *infra* text accompanying notes 27-33.

23. The editor writing the headnote and assigning it to a particular Key Number was engaged in a purposive enterprise — fitting the case into the system. The headnotes were tailored to this purpose. See *infra* text accompanying notes 30-32. Thus, even those merely using the headnotes in a West reporter and not using the Digest proper are affected by the structure.

24. Notice that the West thesaurus is limited to the legal universe. Some universal thesauri are truly universal; they cover the entire universe of subjects. One example is the subject thesaurus of the Library of Congress. SUBJECT CATALOGING DIVISION, LIBRARY OF CONGRESS, LIBRARY OF CONGRESS SUBJECT HEADINGS (9th ed. 1980).

For an excellent summary of indexing theory, see Dabney, *The Curse of Thamus*, 78 LAW. LIB. J. 5, 9-17 (1986).

method of case retrieval.²⁵ But the West system did much more than that. The Key Number System provided a paradigm for thinking about the law itself. Lawyers began to think according to the West categories.

C. Strengths and Weaknesses of the System

The Digest System had both enormous strengths and unresolvable weaknesses. The strengths were the comprehensiveness of the coverage, the reliability and accuracy of the West editorial staff, and the fact that an increasingly large number of published cases could be fit into a recognizable and stable subject format. Furthermore, West was able to maintain the consistency of its system because it was both the publisher and the indexer of cases. These strengths were reinforced by the existence of Shepard's citators, an extremely accurate cross-referencing device.

Perhaps the most important characteristic of the Digest System was that the West editorial staff acted as a national fixed point in the spinning universes of state common law judges and lawyers. The editors were trained to "normalize"²⁶ judicial opinions that used strange language or strange analysis or otherwise appeared to be anomalous, to

25. The West National Reporter and Key Number Systems became an even more powerful research tool when combined with the citation service provided by the Shepard's Company. Shepard's developed its series of citators during the early part of the twentieth century. Frank Shepard was a book salesman who recognized the utility of comprehensive citations for cases. His idea was straightforward: he would provide a service that noted every subsequent mention of a particular decision by any other case. This categorization would allow the researcher ready access to any other decisions which might modify, expand upon, or even comment on the subject decision.

The miracle of Shepard's was its accuracy and comprehensiveness. It covered every court in every jurisdiction. Eventually, Shepard's expanded to cover codes, constitutions and ancillary tools as well. But the heart of the system was always the total, comprehensive coverage of all cases. The other part of this miracle was its reliability. Early on Shepard's established an outstanding record for reliability. The literature of the Shepard's Company on the death of Mr. Shepard demonstrates the company's real pride in the reliable accuracy of its products, and lays out the Company's philosophy in delightfully purple prose:

The present management of Shepard's Citations would have to be men of the dramatic Rabbit type, with hearts of steel and souls devoid of sentiment, to escape the thrills of satisfaction that come with the realization of the worth while [sic] achievements of their organization, and not to realize the enormous debt of gratitude which is owed to the editorial, business, and mechanical forces of the Company for their loyal and unselfish service and their constant devotion to the principles of accuracy which is the one outstanding and indispensable feature from which there must be no departure in any Shepard publication.

PUBLISHERS EDITORIAL STAFF, THE FRANK SHEPARD COMPANY, A RECORD OF FIFTY YEARS OF SPECIALIZING IN A FIELD THAT IS OF FIRST IMPORTANCE TO THE BENCH AND BAR OF THE UNITED STATES 9 (1923).

26. In other words, opinions should be fit appropriately into the West analytic scheme. 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.

bring them back into the orthodox mainstream, to make them fit past cases and present expectations. But the centripetal force exerted on the law by the West staff was also a weakness of the system, as we shall see.

The major weaknesses of the American Digest System were four interrelated problems. The first two problems were quite practical, while the second two were more theoretical. First, the West editors could make mistakes. Second, the tremendous scope of the West universal index combined with the felt need for precision in the subdivisions created a deeply layered index. Third, the universal index was inflexible and resistant to change. Fourth, the editor steeped in the paradigm of the Digest always interpreted cases in a way that fit the paradigm.

1. *Mistakes*

One major weakness of the West System was the fact that the very editor whose job was to "normalize" the judicial language and to correctly index the decision was also subject to human error. The West indexer/editor who wrote the headnote or the individual who assigned the subject location *could* make a mistake. And if this individual misplaced the headnote, it might be lost forever.²⁷

It's difficult to assess the extent of this problem. My impression is that it was not severe. West's reputation for accuracy was well deserved. Still, these kinds of mistakes, even if minor, are eliminated by free-text searching in computer databases, so now they seem like an unnecessary weakness.

2. *Layered Indexing*

There is also a large practical problem inherent in the complex structure of the West Digest. Modern indexing theory criticizes deeply layered indexes. A layered index is one that creates a series of subclassifications in order to increase precision. Many of the topics in the West Digests had such multi-subdivisions, and thus created significant hazards for the searcher. The depth of the indexing in the West Digest System, which resulted from a desire for precision, itself

27. Consider the example of certain purposely "lost" cases. The California Supreme Court occasionally "depublishes" appellate court decisions that have already appeared in the advance sheets by ordering that they not be included in the official reporters. Because West publishes their *California Reporter* and *Pacific Reporter* so rapidly, subsequently depublished cases are sometimes included. However, West does not put the headnotes from these cases into the Digest System. This effectively depublishes the case since no one can ever find it. Nevertheless, lawyers want access to these depublished cases. This has led to their inclusion in the on-line databases. See *infra* note 84.

became a problem.²⁸ A researcher had to figure out not only the first indexing term, but perhaps the second, third and fourth term in order to find the desired case. The West Publishing Company, while striving to provide the best possible descriptive indexing of the case, actually made it harder for the uninitiated to locate items. Only an adept West editor could maneuver with ease through the variegated latticework of sub-sub-sub-classifications. Thus, the Key Number System represented a marvelous achievement in its breadth and precision, but its achievement concealed significant risks for unsophisticated researchers.

3. *The Rigidity of the System*

Another problem of the Digest was the inherent rigidity of the subject structure itself. Naturally, the system developed by Mr. West in the 1880s that supposedly provided a discreet subject category for *every* potential legal issue could hardly have endured for a hundred years without showing some significant strains. But the size of the system argued against active adjustments to it. In order to update the system, the West Company introduced a number of entirely new topics, both at the issuance of each decennial cumulation and during the publication of the General Digest volumes.²⁹ The introduction of such topics required a

28. The example chosen by Dabney, *supra* note 24, at 13, can hardly be improved upon. West "Securities Regulation" Key Number 327 is layered as follows:

- Securities Regulation
 - II. State Regulation (Blue Sky Laws)
 - (C) Offenses and Prosecutions
 - 325. Criminal Prosecutions
 - 327. —Evidence in General

That is five levels of subject breakdown between the user and the lead. And this is just the index!

29. The NINTH DECENNIAL DIGEST, which covered the years 1976-1981, was the first compilation issued after only five years, a response to the growing volume of cases to be processed. This Digest included 24 new or revised topics:

- | | |
|--------------------------------------|------------------------------------|
| Abandoned and Lost Property | Deposits and Escrows |
| Abortion and Birth Control | Dower and Curtesy |
| Accountants | Employers' Liability |
| Administrative Law and Procedure | Extortion and Threats |
| Bankruptcy | Extradition and Detainers |
| Chemical Dependents | Illegitimate Children |
| Condominium | Implied and Constructive Contracts |
| Consumer Credit | Internal Revenue |
| Consumer Protection | Public Utilities |
| Copyrights and Intellectual Property | Urban Railroads |
| Credit Reporting Agencies | Zoning and Planning |
| Debtor and Creditor | |

Herculean purge of the entire system of headnote classification in order to locate all relevant topics and cases and to rearrange them into the new subject order. The difficulty and expensiveness of this process caused West to be fairly reserved in its introduction of topical modifications.

The effect of the natural rigidity of the West System on the legal system is unclear. Nevertheless, it is interesting that American legal literature of the last century was controlled by a paradigm that was naturally both conservative and orthodox during a time when many ascribed these characteristics to the law itself. The West System was conservative in the sense that it resisted change; it was orthodox in the sense that it self-consciously attempted to maintain internal consistency and coherence in American law. The instrument of conservatism was the rigid index. The instrument of orthodoxy was the editorial staff placing new cases into the national index.

4. *The Purposive Role of the Editor*

A trained editor could clarify and "normalize" the language of an opinion. He or she could likewise assess a judge's "real" meaning, read language in context, correct for idiosyncrasies in style or expression and then classify the words in the subject structure. But by intervening in the research process, and by inserting his or her own interpretations, the editor foreclosed other potential classifications of the subject matter. Subsequent researchers always felt the mediating presence of the editor in the very location of the case within the Digest.³⁰

There were two problems that stemmed from the interposition of editorial judgment. First, the West editors had to choose between alternative characterizations of the issues in the case, and to choose between possible locations for the issues in the West subject thesaurus. Disagreements between editors about proper interpretation and classification have been shown to be disconcertingly common.³¹ Even if a certain choice was not a "mistake," it could be less than optimal.

Since 1981 three new or revised topics have been added to the General Digests. They are:

- Children Out-of-Wedlock (1983) (formerly "Bastards")
- Public Utilities (1982)
- Commodity Futures Trading Regulations (1984)

30. Of course, each case is actually inserted into the Digest in a number of locations because the headnotes correspond to the "issues" in the case.

31. See, e.g., Zunde & Dexter, *Indexing Consistency and Quality*, 20 AM. DOC. 259 (1969). Even one editor will classify the same materials differently at different times. *Id.*

In addition to the inevitable indeterminacy of subjective editorial judgments, the judgment of the West editors was inevitably skewed in a particular direction, or, more accurately, frozen in a certain shape. Because of the purposive nature of the editorial process,³² the interpretive range of the West editors was bounded by the intellectual universe of the Digest. Subtle shifts and deflections in the attitudes and language of judges under pressure from new social or legal forces were treated exactly like idiosyncrasy and anomaly. Thus, the greatest strength of the Digest System — its centripetal force, its “normalizing” will to orthodoxy — was also its greatest weakness.

Was the editor, then, a friend or foe? For the pre-computer age lawyer the answer was a resounding “friend.” Because no one could access the cases in any other efficient fashion, even those who despised the digests had to use them. The primary significance of the advent of LEXIS and WESTLAW is that they appeared to eliminate the necessity of a mediating editorial staff, as we shall see.³³

D. The State of Legal Literature, B.C. (Before Computer)

For all of its flaws, the complete set of case reports with its comprehensive subject arrangements were powerful tools. By the middle of the twentieth century, the legal system had available to it in the publication of its cases a comprehensive system of document production. This system included total subject availability and a comprehensive and accurate system of citation. It made legal literature unique among other disciplines.³⁴ No other discipline had invested the resources and time to develop these extremely efficient manual systems.

Thus, when on-line databases first appeared, they were not particularly attractive to the legal researcher. The possibility of comprehensive retrieval and indexing combined with the citation services offered on the new databases were no nirvana for the legal researcher. She already had such tools available on her desk. There was no need to wait for the

32. See *supra* note 23.

33. See *infra* text accompanying notes 35-47.

34. The relatively high degree of integrity and cohesiveness in legal literature is probably inevitable, with or without the historical fact of the West System, because law is a field where the primary source materials have normative force. In most disciplines, researchers are interested in sources because of the quality the work and the intellectual power of the authors (or the lack of it), or out of historical interest; but later workers in such disciplines are in no sense *bound* by the work of their predecessors. In legal literature, the primary materials (cases) provide (or, for the legal realist, appear to provide) legal workers with a crucial type of binding social norm — law — that people have to know in order to structure their relations with others, and to restructure relations that have broken down.

conversion of the information or to suffer through the inevitable training and start up difficulties. As a result, law was slow to turn to on-line sources for information.

But the change did come.

II. THE NEW PARADIGM: THE ADVENT OF LEXIS AND WESTLAW

A. The Development of the System

The LEXIS system appeared nationally in the mid-1970s. The enormous cost of creating a full-text database of cases was a significant entry barrier, but Mead Data Central persevered and, with the active cooperation of state bar associations, eventually expanded its LEXIS database to include judicial opinions from every state. The history of this expansion is described elsewhere.³⁵ The end-product was a national system containing on-line the full text of every printed case.³⁶

As LEXIS made headway, West introduced WESTLAW. In its first incarnation WESTLAW did not contain the full text of court decisions. Instead it utilized only the text of the headnotes. WESTLAW used the power of the computer and free-text searching to enhance its already existing manual system. This decision proved to be a disaster. West had failed to grasp the nature of the new research tool, and the real significance of the new form of legal literature. Why would a lawyer bother to learn the mechanics of computer research to access the Digest System which had been designed and perfected as a manual, hard copy research tool? West soon caught on and began including the full-text of decisions in addition to the headnotes.³⁷

When LEXIS initially marketed its system, the most frequently heard criticism was that free-text Boolean searches were inappropriate for retrieving judicial opinions. Critics pointed to the variety in judicial language and to the difficulty of locating desired opinions by attempting to specify exact common terms. For example, in 1975, Professor J. Myron Jacobstein of Stanford Law School challenged LEXIS demonstrators at an American Association of Law Libraries convention. He described the facts and law of a particular case concerning a child and asked them to locate it. What Professor Jacobstein knew and the demonstrators did not was that the opinion had uniformly referred to

35. See Harrington, *A Brief History of Computer-Assisted Legal Research*, 77 L. LIBR. J. 548 (1985); Burson, *Report from the Electronic Trenches: An Update on Computer-Assisted Legal Research*, LEGAL REFER. SERV. Q., Summer 1984, at 3.

36. Actually, the full-text databases generally extend their coverage only back to the 1920s or 1930s, although backward expansion continues.

37. This is the "Full-Text Plus" system described in note 68, *infra*.

the child as an "infant." Because both LEXIS and WESTLAW could only retrieve the exact terms entered into their databases,³⁸ the chagrined demonstrators could not find the case.

Both systems struggled to resolve these problems. LEXIS and WESTLAW added the capability to truncate search terms so that the searcher could retrieve all items by searching with the roots of words.³⁹ In addition, both systems altered their search software to automatically retrieve plurals and convert statutory alphanumerics.⁴⁰ As familiarity with the systems grew, researchers waxed in confidence, and the relevance of the kinds of questions posed by the Jacobstein challenge seemed to fade.⁴¹

B. The Structure of the System: Full-Text Databases and Free-Text Searching

LEXIS and WESTLAW's use of a full-text format was a big step. The computerized research systems that were coming into use in other disciplines generally did not contain the full text of documents. Instead, they consisted of abstracts or index entries.⁴² These systems were highly

38. If, for example, the searcher wanted to find all cases that analyzed the rights of unwed fathers concerning adoption of their children, she could frame the search as:

father & child w/15 adoption

This search strategy would retrieve all cases that contained both the word "father" and the word "child," where the word child appeared within fifteen words of "adoption." The Jacobstein challenge demonstrated that if a judge had referred to the child throughout the opinion as "infant" or "son" or "baby" or "minor" the case would not be retrieved.

As we shall see, the searcher can attempt to resolve this problem by including synonyms in the search request, but this search strategy results in the retrieval of an unwieldy number of cases. Many of the cases found by such a search will be irrelevant. As the number of search terms are increased, the retrieval of unwanted items escalates. The searcher is therefore placed in a dilemma. To ensure the retrieval of desired items she must expand the search. Such action, however, increases the search's cost while it likewise increases the inefficiency of the retrieval. This is the problem of the inverse relation between Recall and Precision. See *infra* notes 51-56 and accompanying text.

39. For example, the search described in note 38 could be changed to:

father & child w/15 adopt!

This search would retrieve cases that used words like "adopts" or "adopting" in addition to "adoption."

40. There is speculation that the search software will be modified by so-called "artificial intelligence" techniques to include synonym retrieval, but that does not seem likely in the near future.

41. However, several recent papers have raised it again in sharp relief. See *infra* text accompanying notes 48-72.

42. Examples are MEDLINE, a database of medical information that indexes and provides bibliographic citations to articles in over 3,000 journals and chapters from selected monographs, and SOCIAL SCISEARCH, which indexes and provides bibliographic citations to articles in 4,500 social science and scientific journals. Neither offer the full text of the indexed documents.

efficient search tools, but they left the researcher with the task of document location. In addition, they retained all of the problems associated with the mediating role of the indexer. LEXIS and WESTLAW, on the other hand, were based on a different concept — the idea that the researcher needed a totally integrated system that freed him from any index-imposed restraints and that allowed him to examine the full document on-line. The full-text feature made the systems more expensive than on-line index systems,⁴³ but they also provided the innovation that attracted users.

43. Also, the expense of the systems was added to that of traditional research tools. The current cost of the systems is hard to assess and compare because WESTLAW charges a regressive flat rate for on-line time (starting at \$150 per hour for the first three hours per month) while LEXIS charges a fixed amount per file access, search, search modification, plus on-line time.

The law librarian at Control Data Corporation recently published some comparisons that are useful. Griffith, *Dual-System Research: The Best of Both Worlds*, Legal Times, March 17, 1986, at 9, col. 1. Griffith divided the searching universe into four categories that roughly corresponded with search habits in his office, and ran identical searches on each system: (1) five "search and browse" searches — researching more than 10 minutes on-line; (2) 12 "search and cite" searches — quick-answer research (less than 10 minutes on-line); (3) eight "cite check or retrieve" searches — using Shepard's or retrieving a specific case; and (4) five "case retrieval" searches — finding and retrieving a case of unknown citation. His findings were:

Research Request Type and No.	WESTLAW Retrievals/Cost	LEXIS Retrievals/Cost
(1) "browse" / 5	98 / \$ 127.19	108 / \$115.31
(2) "quick" / 12	1202 / \$ 57.89	1616 / \$207.31
(3) "cite" / 8	— / \$ 17.01	— / \$ 13.60
(4) "find" / 5	— / \$ 9.66	— / \$ 62.58

These search costs are significant, but notice that both systems give users access to (at a minimum) all cases in the West National Reporter System. Bound volumes of all these cases are not cheap. The base prices for West reporter sets are:

Reporter Set	Volumes	Price
Atlantic Reporter 2d	1-499	\$ 12,081.75
Federal Reporter 2d	1-776	\$ 14,537.50
Federal Supplement	1-617	\$ 11,876.00
Northeastern Reporter 2d	1-484	\$ 11,746.75
Northwestern Reporter 2d	1-375	\$ 8,502.50
Pacific Reporter 2d	1-706	\$ 16,732.75
Southeastern Reporter 2d	1-335	\$ 7,357.50
Southern Reporter	1-477	\$ 11,744.50
Southwestern Reporter	1-697	\$ 14,026.25
Supreme Court Reporter	1-106	\$ 3,308.50
Total		\$ 111,914.00

(Additional volumes cost about \$35.00, and each additional term of the United States Supreme Court costs \$137.50.).

Prices obtained from Donald Blockhus, Sales Rep., West Publishing Co. (Mar. 1986) (available at *High Technology Law Journal*). Obviously, a set of reporters would pay for a lot of computer time. For a law office that mainly uses cases from a few jurisdictions, it

With LEXIS and WESTLAW the researcher could locate material by using, not a predetermined subject thesaurus or index,⁴⁴ but the free-text searching method. Using search commands that incorporate Boolean logic, the researcher retrieved documents by requesting cases that contained a specific term or terms. Boolean logic allowed the terms to be linked by occurrence, proximity, section of a document and various combinations thereof.

The real breakthrough with LEXIS and WESTLAW, however, is that they eliminate the intervention of *any* editorial judgment. It is now possible to research efficiently without the mediating presence of the West editors. No editor or index stands between the language of the opinion and the researcher as he or she frames the search request. This simple fact vaporizes the full range of complaints that had accumulated against the old West system. Editors and antiquated subject structures no longer burden the research process.

C. Strengths and Weaknesses of the System

The full-text, free-text searching of the on-line literature frees researchers from many of the serious flaws of the old paradigm. First, an editor can no longer "misplace" a case by misinterpreting a decision and placing it in the index in a way that forecloses access by research-

may be cheaper to use a computer system to do research in other jurisdictions than to buy a set of seldom used reporters. Of course from a client's point of view, these price differences may not be significant because the cost of a lawyer's time dwarfs the costs of any legal research method.

44. To increase retrieval speed, full-text databases actually have an index, but the index contains every word and word root in the database, along with a description of every location of that word in the database. (Words like "a" or "the" are not included in this list.). This kind of index is called a "concordance." For example, the word root "adopt" is indexed along with each of its locations by document, paragraph, sentence and position in the document. The words "father" and "child" are indexed with the same location information. If the search request is:

father & child w/15 adopt!

the computer will find "child," "father," and "adopt" in the concordance and compare all of their locations. Whenever "child" and "adopt" occur within 15 words of each other in a document that also contains "father" the computer will retrieve that document.

The concordance scheme is essential. If the computer had to search every document in the enormous database one by one, the searches would be interminable. Also, users of the systems searching for some word "x" may have noticed that they receive the message "The word 'x' is not in the database" extremely rapidly, far faster than a search that actually retrieves documents. This is puzzling until the user understands that the first part of the search is a search of an index that contains every significant word in the database.

ers.⁴⁵ Every case is equally available to every researcher, limited only by the researcher's training and ingenuity.

Second, there are a number of functions that researchers with full-text database capability can perform that could have been performed only inefficiently or not at all by a person employing manual research techniques. Intelligent use of the "segment searchers" on both LEXIS and WESTLAW systems is one good example; cross referencing search terms by a particular judge, a particular court, a particular date, or even by such indicia as a name of a particular party can yield helpful and practical information. Or, when legal research problems concern a specific object, one that has a unique name or phrase that describes it — for example, a product or trademark — the computer searches can yield every use of the unique term in the entire corpus of cases. Also, a researcher can find every case mentioning a certain code section or a certain previous case. This kind of research was simply impossible in the old system, and it can be extremely useful.

Third, the new form of legal literature eliminates the rigidity inherent in the West paradigm. In the old system, the information in each of the cases was parsed into a preexisting framework that inevitably tended to suppress subtle changes and to enforce judicial and professional conformity and conservatism. In the new legal literature, the information is strewn into a free-form database without differentiation. The legal databases provide no guidance and place no restrictions on the way that judges and lawyers think about cases. The use of specific words and the presence of specific facts become more important to the researcher than the "holding" of a case or any other abstract generalization about the law. The new paradigm is not merely a more flexible structure than the old. The new paradigm has no structure at all.⁴⁶

Fourth, and perhaps most important, the absence of an index means the absence of indexers. There is no "normalizing" editorial force, no will to consistency, coherency or orthodoxy. To the extent that these characteristics are seen as desirable, the responsibility for maintaining them is placed squarely on the shoulders of judges and lawyers, and not on an anonymous functionary in the bowels of West Publishing.

45. Sometimes, however, typographical errors in the databases can have a similar effect. See *infra* note 65.

46. The databases have no subject-matter structure. The division of the legal databases into "files" or "libraries" provides a sort of structure, but this generally amounts to classification by jurisdiction, precisely the arbitrary system of classification abjured by the West National Reporter System and American Digest. However, information pertaining to certain specialized areas of practice, e.g., trade regulation and bankruptcy, are increasingly gathered together in special files and libraries in both systems.

The legal databases make available the raw materials of legal research as never before: raw.

But, inevitably, the new system has created its own problems, problems inherent in the new research process and the new form of legal literature.⁴⁷ The two most significant practical problems are, first, the questionable efficacy of free-text computer searching, particularly in enormously large databases, and, second, the tendency of those using and promoting the new paradigm to see every lawyer as an appropriate end-user of the systems. The third problem is more theoretical: what kind of legal practice will cohere with a form of legal literature that makes judicial opinions available according to practical search skills and that interposes no mediating and integrating editorial judgment between the raw legal materials and the practitioner?

1. *The Mechanical Limits of the System: the Efficacy of Free-text Searching*

The Jacobstein challenge demonstrates that lawyers have shown some concern about the efficiency and accuracy of free-text searching since the inception of the legal databases. These concerns have focused on the ability of free-text searching to deal with the vagaries and variety of thought in language, and the concern with the effect of even small errors in the huge legal databases. I will examine these criticisms by reviewing two excellent articles that have brought them back into the limelight.

- a. The Reemergence of the Issue: Blair and Maron's Study

In March, 1985, David Blair and M.E. Maron published an article that caused a flurry of interest among legal researchers.⁴⁸ The two researchers had a marvelous opportunity. They worked with a large, operational, full-text document-retrieval system that was set up to serve as a litigation support system in an actual case. The system contained approximately 40,000 documents (roughly 350,000 pages of text) that were thought pertinent to the defense of the lawsuit. With complete access to a large full-text database, with search software similar to LEXIS'

47. In many respects, the problems of the new paradigm appear to be the flip-side of the problems of the old paradigm. This fact, along with conventional prudence, seems to suggest that currently the optimal research tool is the two systems used together. See *infra* text accompanying notes 87-89.

48. Blair & Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System*, 28 COM. ACM 289 (1985).

and WESTLAW's,⁴⁹ and with sufficient funding to back them,⁵⁰ Blair and Maron were in an unusual position to attempt a test of the efficiency of full-text search systems.

Blair and Maron were primarily interested in two measures of retrieval effectiveness: Recall and Precision. "Recall measures how well a system retrieves *all* the relevant documents; and Precision, how well the system retrieves *only* the relevant documents."⁵¹ If Recall is low the system is retrieving only a small percentage of the total number of relevant documents. If Precision is low the system is retrieving too many useless documents. In full-text searching systems, Recall is inversely related to Precision.⁵² Most lawyers probably would be more immediately concerned with Recall. They would want all the relevant materials, even if they have to weed out a lot of irrelevant stuff. But in reality, Precision is just as important. Low Precision in a large database produces what researchers call "output overload." A high Recall/low Precision search in a large database might retrieve 1000 documents of which 700 or 800 are irrelevant. Most organizations don't have the necessary time or resources to cull that much information.

The study utilized a database searching team made up of two legal assistants and two attorneys, all of whom were intimately familiar with the case and the content of the computerized litigation file. When an attorney wanted to see certain information from the file, he or she would give a written description of the research to one of the assistants. The legal assistant would frame an inquiry and run a computer search. The results of the search were evaluated by the requesting attorney.⁵³ If an attorney was not satisfied that 75% of the relevant documents in the database had been retrieved, he or she would ask to have the query reformulated and run again. The research was considered complete (usually after a number of searches) only when the attorney was satisfied that the search had produced 75% of the desired documents. When the attorney was satisfied, Blair and Maron's team would compare the number

49. The search software was IBM's STAIRS, an acronym for STorage And Information Retrieval System. *Id.* at 289.

50. Their project cost almost half a million dollars in direct and indirect expenses. *Id.* at 298.

51. *Id.* at 290. Recall is the ratio of the relevant documents retrieved by the search to the total number of relevant documents in the database. For example, if a database consisted of 1000 documents, 100 of which were relevant, then a search that retrieved 50 of the relevant documents would have 50% Recall. Precision, on the other hand, is the ratio of relevant documents retrieved to total documents retrieved. For example, if a search retrieved a total of 75 documents, 50 of which were relevant, then the Precision of the search would be $50 \div 75 = 66\%$.

52. *Id.* at 293.

53. *Id.* at 291.

of relevant documents retrieved by the search (or searches) with the total number of documents retrieved to determine the Precision of the searches.⁵⁴ The computation of Recall was much more complex,⁵⁵ but it amounted to a very conservative estimate of Recall.

The results were surprising and dismaying. The full-text retrieval litigation support system proved to be a fairly inefficient search mechanism. On the average it retrieved about 20% of the desired documents, i.e., Recall was about 20%. On the other hand, Precision was relatively high at about 79%. The study also confirmed the observation of earlier studies that Recall and Precision are inversely related.⁵⁶ Even more interesting was the fact that the lawyers working with the research team had estimated the Recall efficiency of the system at a minimum of 75%.⁵⁷

The most crucial fact about the Blair and Maron study is that it was the first time a file of this size was used to study full-text searching with Boolean operators. Seminal studies that "demonstrated" the desirability of full-text searching had used smaller databases.⁵⁸ It was only because

54. *Id.*

55. Because it was impossible to have the two attorneys (who were making all relevancy determinations) read the entire 350,000 pages of text in order to find all of the relevant items, the researchers had to find another way to calculate Recall. However, it is not clear from the article how Recall was estimated. Their explanation is contained in one paragraph:

To find the *unretrieved* relevant documents, we developed sample frames consisting of subsets of the unretrieved database that we believed to be rich in relevant documents (and from which duplicates of retrieved relevant documents had been excluded). Random samples were taken from these subsets, and the samples were examined by the lawyers in a blind evaluation; the lawyers were not aware they were evaluating sample sets rather than retrieved sets they had personally generated. The total number of relevant documents that existed in these subsets could then be estimated. We sampled from subsets of the database rather than the entire database because, for most queries, the percentage of relevant documents in the database was less than 2 percent, making it almost impossible to have both manageable sample sizes and a high level of confidence in the resulting Recall estimates. Of course, no extrapolation to the entire database could be made from these Recall calculations. Nonetheless, the estimation of the number of relevant unretrieved documents in the subsets did give us a *maximum* value for Recall for each request.

Id. at 291-92 (emphasis in original). It's hard to see how the last sentence — claiming a maximum value for Recall — follows from the explanation.

In a telephone conversation, author M.E. Maron explained the Recall calculation as follows. The authors found rich subsets by using very broad search techniques. (For a more complete description of this process, see Dabney, *supra* note 24, at 28-29.) They took random samples from these rich subsets and had the lawyers evaluate them thinking that they were search results. From the number of relevant documents in the random samples, they extrapolated Recall for the rich subsets. Then, they postulated that there were *no other* relevant documents in the database. This technique, although still an estimate, is quite conservative and does approximate a *maximum* value for Recall because there were bound to be relevant documents outside the rich subsets.

56. *Id.* at 293.

57. *Id.* at 295.

58. *E.g.*, Salton, *Automatic Text Analysis*, 168 *SCIENCE* 335 (1970); Swanson, *Searching Natural Language Text by Computer*, 132 *SCIENCE* 1099 (1960).

of an unusual opportunity that Blair and Maron could afford the extended time and effort needed to evaluate a large system. Their evaluation raises some serious questions.⁵⁹

b. The Curse of Thamus: Finding Words but not Wisdom

One explanation for the very low recall rate described by the Blair and Maron study is that the human use of language is inexact. Full-text searching is premised on the assumption that "it is a simple matter for users to foresee the exact words and phrases that will be used in the documents they will find useful"⁶⁰ The problems of imprecise usage, synonyms, jargon and even misspellings challenge this assumption. To quote Blair and Maron, "it is impossibly difficult for users to predict the exact words, word combinations, and phrases that were used by *all* (or most) relevant documents"⁶¹ Daniel Dabney provides an interesting analysis of these kinds of problems in his recent article *The Curse of Thamus*.⁶²

Dabney divides the problem of matching words into three categories: synonymous words, ambiguous words, and complex expressions. The first two categories — synonymous words and ambiguous words⁶³— involve the problem of linguistic imprecision. Because judges can refer to a person or a thing in many different ways, it is difficult to be certain that any search term or terms will retrieve the relevant cases.

Blair and Maron argue that if the earlier studies had utilized large databases they would have reached less sanguine conclusions. Unlike the litigation database investigated by Blair and Maron, the small databases could be searched with high Recall and low Precision techniques without "output overload." Blair & Maron, *supra* note 48, at 298.

59. There is at least one potential limitation on the applicability of the Blair and Maron study to the LEXIS and WESTLAW systems. A litigation support file contains a heterogeneous mix of documents that includes, among other items, reports, memos, letters, invoices, transcripts of meetings, conversations, etc. The LEXIS and WESTLAW databases are primarily composed of judicial opinions, a relatively homogeneous form of discourse. To use Dabney's example, a judge might call a child a "minor" or an "infant," but it is unlikely that he will call a child a "punk" or "rug rat." Still, even if free-text searching in an on-line legal database were *twice* as efficient as the litigation support database studied by Blair and Maron, a 40% recall rate still would be uncomfortably low.

60. Blair & Maron, *supra* note 48, at 295.

61. *Id.* at 295. This is, in substance, the same concern raised by critics at the advent of the legal databases. See *supra* text accompanying notes 37-40.

62. Dabney, *supra* note 24. The title of the article comes from a legend in the Phaedrus of Plato. According to Plato, the Egyptian King Thamus disapproved the invention of writing by the god Theuth. Thamus thought that because the mere possession of writing could not give wisdom, writing would cause far more harm than good. Dabney notes that we possess an almost unimaginable amount of writings, but asks "how are we to extract from this almost incomprehensibly large collection of written records the knowledge that we need?" *Id.* at 5-6.

63. *Id.* at 18-19.

Dabney illustrates the problem of synonyms by giving the example of a search for a case concerning a ten-year-old boy. The court might refer to the boy as "boy," "minor," "child," "juvenile," "youth," "ten-year-old," "infant," or "young man."⁶⁴

Ambiguous words create the converse problem. The searcher may use an apparently specific word that has few or no synonyms and that should isolate the relevant cases, only to find that the word has an entirely different meaning. Dabney's example is a researcher looking for cases involving the drug DES (diethylstilbestrol), and retrieving *Tinker v. Des Moines Independent Community School District*.⁶⁵ This problem is augmented by the ability to search for word roots. For example, in a search for cases involving the adoption of a child, the searcher might attempt to retrieve cases that use the noun "adoption," the verbs "adopted" and "adopts," and the adjective "adopted" by using the following search:

father & child w/15 adopt!

This search would also retrieve cases involving a father and child were the opinion "adopts" a rule of law or a particular version of a disputed factual finding.

64. *Id.* at 18. Blair and Maron provide an amazing example of the problem of synonyms.

Sometimes we followed a trail of linguistic creativity through the database. In searching for documents discussing "trap correction" (one of the key phrases), we discovered that relevant, unretrieved documents had discussed the same issue but referred to it as the "wire warp." Continuing our search, we found that in still other documents trap correction was referred to in a third and novel way: the "shunt correction system." Finally, we discovered the inventor of this system was a man named "Coxwell" which directed us to some documents he had authored, only he referred to the system as the "Roman circle method." Using the Roman circle method in a query directed us to still more relevant documents, but this was not the end either. Further searching revealed that the system had been tested in another city, and all documents germane to those tests referred to the system as the "air truck." At this point the search ended, having consumed over an entire 40-hour week of on-line searching, but there is no reason to believe that we had reached the end of the trail; we simply ran out of time.

Blair & Maron, *supra* note 48, at 295. Of course, this example comes from a heterogeneous litigation support file, not from a set of judicial opinions. Still, anyone with a little imagination can think of a similar trail through the cases.

65. 393 U.S. 503 (1969). This example is somewhat deceptive. A court that used the abbreviation DES is likely (though not certain) to have used the full term at some point. For example, I searched for the term "des" in the LEXIS States/Omni database on March 2, 1986, and retrieved 6888 cases. I browsed through the first thirty cases and found that nearly all involved either the city Des Moines or an Alaska criminal case, *Des Jardins v. State*, 551 P.2d 181 (Alaska 1976). I then searched for the term "diethylstilbestrol" and retrieved 62 cases and 33 ALR annotations, a much more manageable search. Of course, the second search may have been incomplete.

My "des" search also brought home the thorny problem of typographical errors in the databases. Three of the first 30 cases were "hits" because of typographical errors, including two misspellings of "does," and one misspelling of "describe." This rate may not be representative, but it is nevertheless disconcerting.

But the problem is larger than the mere "imprecision" of language — for example, whether a child will be called an "infant" or a "minor." The fact is that law involves ideas, and ideas are not directly correlated with particular words.⁶⁶ Dabney describes this as the problem of complex expressions, his third analytical category.⁶⁷

The difficulty of matching words with ideas is in some ways more insurmountable than the problems of matching words with persons or things. On the one hand, research problems that involve specific factual questions or specific statutes or administrative rules are quite amenable to straightforward computerized searches. Also, the *skillful* researcher can develop strategies for searching with words that have specific denotations but synonyms or multiple meanings. But for searches involving legal concepts — or any ideas that can be expressed without using a particular word or phrase — the computers are not very effective. Conceptual questions are difficult to frame in the Boolean search strategy because judges are not likely to use exactly the same words to describe the same ideas or concepts.⁶⁸

66. See Childress, *supra* note 8, at 1533:

Time cannot correct the inherent limitations of the word-search method, however, and concordance logic may produce its own inefficiencies. LEXIS' dependence on words, for example, grounds search capabilities in the opinion's language rather than its content. An unusual or incomplete description of the facts or issue may "lose" a very relevant case from a reasonable search.

67. Dabney, *supra* note 24, at 19.

68. West Publishing attempts to solve this problem with its "Full-Text Plus" system. Full-Text Plus refers to the fact that the WESTLAW database contains the full text of cases *plus* the same text of headnotes and Digest summaries printed in the National Reporter System. West claims that this addition introduces "normalized" language because the trained editor has again entered the picture. The uniform language in the headnote and syllabus are supposed to compensate for the imprecision of the judicial author. Thus, the searcher can formulate a search strategy knowing that his search phrase will be matched up both with the text of the judicial opinion *and* with the "normalized" language introduced by West editors in the headnotes and case synopsis.

A recent study by Professor Al Coco lends some credence to this claim. Coco, *Full-Text vs. Full-Text Plus Editorial Additions: Comparative Retrieval Effectiveness of the Lexis and Westlaw Systems*, LEGAL REFER. SERV. Q., Summer 1984, at 27. The study indicates a substantial difference in retrieval produced by running the same search on both LEXIS and WESTLAW, with the latter consistently retrieving more cases.

Dabney has questioned this result, noting that no relevancy verification of the cases was made. He has also questioned the basic theory that Full-Text Plus' addition of headnote and synopsis language is a major amelioration of the problem. Dabney's point is threefold: (1) headnote language invariably tracks the text of the case, thus adding little in the way of "normalized" language; (2) while subject headings accompany the headnote in the database, only two levels (the highest and lowest) of West's deeply-layered subject structure are included, and therefore, most of the relevant headings are dropped; and (3) because the synopsis paragraph is so general, it is of marginal assistance to the searcher. Dabney, *supra* note 24, at 31-34. Also, my impression is that WESTLAW edits its inputted material more carefully than LEXIS, so that the additional "hits" found by Coco may have resulted from correct spelling as well as from Full-Text Plus. See *supra* note 65.

Dabney summarizes this point with an excellent example. He postulates a search for the question:

"If a person waives his or her right to trial by jury in one trial, can a jury trial still be demanded in a subsequent new trial of the same matter?" The key words for this question, "trial," "jury," "waiver," and "retrial" are common in judicial opinions, but discussions of the specific point of law of the question are relatively rare. A computer cannot reliably find cases that are on point because too much of the meaning of the desired cases is tied up in the syntactical relationships between the words, which are not "understood" by the computer.⁶⁹

In other words, unless a particular legal concept can be reliably mapped to a relatively unique word or set of words, the concept will be invisible to the researcher on a free-text system.⁷⁰

The Blair and Maron study indicates that the problems outlined by Dabney remain. Indeed, as the size of the databases expands, so does the magnitude of the problem. The body of case law increases dramatically each year. The West Publishing Company calculates that it adds 65,000 full opinions to the corpus annually.⁷¹ These numbers give an idea of the truly monstrous scope of the legal databases. And this becomes a problem in itself.

c. Error Rates and the Staggering Size of the Databases

The sheer size of the databases is a primary source of inefficiency. Retrieving 10% of a database of 100 documents presents few problems. The ten documents that contain search terms used in the research query will yield a manageable file that can be scanned easily to assess relevance. If the file contained 40,000 documents, however, a 10% retrieval rate would produce 4,000 documents. A researcher cannot thoroughly evaluate such a large number of documents. In fact, 30 documents may be too many. The researcher needs to construct high Precision search strategies that recover mostly relevant documents; unfor-

69. Dabney, *supra* note 24, at 19-20.

70. An excellent example of this difficulty is LEXIS and WESTLAW's failure to market effectively on-line databases of state statutes. Because the content of statutory materials is highly conceptual and uses language that is either repetitive or *sui generis*, it is relatively inefficient to research with free-text searching. Also, these materials are costly to load on-line. Although WESTLAW is experimentally loading Illinois statutory material, neither LEXIS nor WESTLAW plan in the future to market state statutes, and, in part, their decision originates from these problems.

71. West supplied this figure as a part of a packet of information distributed during a Summer 1985 tour. The estimate was confirmed by Bill Lindberg, a West administrator, in a telephone conversation.

tunately, this strategy is bound to exclude many relevant documents as well.

Dabney analyzes this problem in detail. He explains the inevitable dilemma in which the researcher is caught. As the searcher expands the search to retrieve all relevant cases, he or she pulls in irrelevant ones as well. In order to screen out irrelevant materials, the searcher will add more detail to the search request. This strategy does reduce the number of cases retrieved, but it also contributes to the exclusion of relevant materials.⁷² This is the problem of the inverse relation between Recall and Precision described by Blair and Maron. Their study demonstrates that when the researcher conjoins additional search terms to reduce the size of the search output, more and more relevant documents are excluded. As the LEXIS and WESTLAW databases continue to expand in size these difficulties will only be exacerbated.

2. *The Limits of the User*

The second basic problem with computer-based free-text searching is the limitations of the individuals who use the computer. Given the limitations of free-text searching, who should be expected to search an on-line, full-text database effectively and to evaluate the quality of his or her search? In the language of information science, who is the proper end-user?

Well-trained and experienced computer searching experts are more effective full-text computer searchers than subject-matter experts.⁷³ But in the legal profession, most LEXIS and WESTLAW searching is conducted by lawyers. Both legal database sellers push the model of a law office with a database terminal on each lawyer's desk. Are lawyers the proper end-users of the full-text databases?

a. *Training Incompetents . . . or Worse*

The first issue is the adequacy of lawyers' training. Although both database vendors make their own training systems available to law firms who subscribe, most attorneys are first exposed to and receive their basic training on the systems during law school. Every accredited law school in the United States now has either a LEXIS or WESTLAW terminal, and an increasing number have both. The task of training students in the use of these on-line systems has become their responsibility.

72. Dabney, *supra* note 24, at 21-26.

73. Cf. Curry, *The Value of the Search Request Form in the Negotiation Process Between Requester and Librarian*, 20 AM. SOC'Y INFO. SCI. PROC. 115 (1983); Obermeier, *Expert Systems — Enhancement of Productivity?*, 20 AM. SOC'Y INFO. SCI. PROC. 9 (1983).

Unfortunately, most law schools have spotty records for *any* kind of research training. The discussion in the literature on the failure of manual research training programs is vast.⁷⁴ Most law schools have made little headway in solving the age old problem of how to train their students in traditional research methods. With this unstable foundation it is unlikely that law schools will successfully handle their new responsibility for training efficient on-line researchers.

In many law schools computer training is the responsibility of the law library staff. Very rarely does the staff at these law libraries receive enough money to develop a truly successful training program. Other law schools hire students who are already familiar with computers and/or LEXIS and WESTLAW to train students. At a conference held during the summer of 1985, a group of people who were brought together because of their expertise in providing LEXIS and WESTLAW training admitted that their own programs did not adequately train potential users. This group concluded that the most that could be asked of a law school training program is that it acquaint the computer user with the capacities of the system.⁷⁵ Due to the skewed ratio of trainers to students and number of machines to students currently involved in LEXIS and WESTLAW training, it is not possible to train each student to be an effective and efficient searcher.⁷⁶

74. A personal favorite is Brock, *The Legal Research Problem*, 24 DE PAUL L. REV. 827 (1975). See also Mills, *Legal Research Instruction in Law Schools, the State of the Art, or, Why Law School Graduates Do Not Know How to Find the Law*, 70 L. LIBR. J. 343 (1977); Achtenberg, *Legal Writing and Research: The Neglected Orphan of the First Year*, 29 U. MIAMI L. REV. 218 (1975).

75. These observations are based on a discussion at a West Publishing Conference held in August 1985, at St. Paul, Minnesota. Both West and Mead Data have instituted regional workshops for law librarians to discuss the problems training law students in the use of their systems and possible solutions.

76. Both LEXIS and WESTLAW are now concentrating on assisting law schools in introducing their students to the on-line systems. During the summer of 1985, both systems sponsored special workshops to talk to legal educators about such training programs. Moreover, in order to train students one-on-one, both systems have made available to large law schools a number of terminals on a temporary basis. LEXIS and WESTLAW experimented with these Temporary Learning Centers during the 1985-1986 school year. In some locations LEXIS and WESTLAW are setting up Permanent Learning Centers (PLCs) in law schools. PLCs allow LEXIS and WESTLAW to train law firm subscribers at a local law school library. When the database vendors are not using these terminals for their own professional training programs, the law schools are free to use them.

Recently, both LEXIS and WESTLAW have developed another training program. During the summer of 1985, each announced that it would allow law school subscribers to use the schools' terminal subscriptions as a free route of entry for up to three personal computer users of the same system. This will permit the Deluxe terminal and three personal computers to be in use at the same time as a part of the same subscription. The only limitation is that the usage be confined to off-peak hours. But despite these

Even if students originally were trained in the efficient use of LEXIS and WESTLAW, including requisite skepticism about the usefulness of free-text searching, the frequent changes in the databases and the constant stream of enhancements call for continuous *retraining*. Few lawyers can commit the time or energy to maintain their skills. This means that, at best, law schools are graduating students who think that they have been trained in the use of LEXIS and WESTLAW, but who approach the systems with little or dated sophistication. The ease of use of the legal databases may actually compound these problems by giving lawyers a false sense of competency.

b. User-Friendly or User-Seductive?: The Moron Cadillac

When the Mead Data Central Company first marketed LEXIS, it bundled access to LEXIS with a dedicated terminal.⁷⁷ The LEXIS "Deluxe" dedicated terminal was large and ugly, but its operation was a model of simplicity. The distinctive labeling of the keys allowed even the most unsophisticated user to quickly master the mechanical aspects of terminal operation and interaction with the search software. For example, if a lawyer wanted to see the next case, she simply pushed the button labeled "next case."

For years I have described the LEXIS Deluxe terminal as a "moron Cadillac," designed so that it could easily be operated by even the most machine-resistant lawyer. The premise of the design (a correct premise I might add) was that the average practicing lawyer would not read accompanying "How to Use" manuals, nor would he or she attend training sessions. Lawyers intimidated by computer jargon and worried about interacting with computer trainers found that they could almost "train" themselves. After a false start, West eventually introduced a similar terminal with dedicated keys.⁷⁸

efforts by the vendors, the burden of training still lies with law schools, and it is not clear that they can shoulder it effectively.

77. A dedicated terminal is one that is designed to be used in a particular application; generally, it is not compatible with other systems.

78. The original terminal marketed by the West Publishing Company as part of its WESTLAW on-line system was a more or less standard dumb terminal, not nearly as user-friendly as the LEXIS Deluxe. It required a higher degree of computer sophistication because it required that the operator learn and understand command codes. On the other hand, because the terminal was not dedicated, it could be used to access other databases and computer systems. West thought that this innovation would be a significant economic advantage to the user. The West terminal also cost much less than the moron Cadillac. Because of the premium placed on price and flexibility in most parts of the information industry, West's marketing decision appeared sound. The LEXIS Deluxe terminal was expensive and could not be adapted for other uses. But the legal community as consumer made its own judgment, and the LEXIS Deluxe terminal was a great success.

The simplicity of the user friendly terminals is more problematic than it might first appear. By encouraging the lawyer to believe that he has the requisite sophistication to use the system, these terminals may delude the researcher into overestimating his or her abilities to search effectively. The ability to operate the terminal and to sort through libraries and files does not guarantee adequate searching skills. The simplicity of the terminal's operation permits a lawyer to attend a training session and then to allow his skills to atrophy because it is months before he plops down in front of the terminal again. At that point he will be able to puzzle out the mechanics of operation, but that is no guarantee of effective searching. This is one significant source of inefficient and expensive searches.

As both LEXIS and WESTLAW make their systems available for use with personal computers, the problem of inefficient searching both deepens and widens. The problem deepens because individuals who operate their own personal computers to search the LEXIS or WESTLAW databases may be used to other on-line databases, which generally use traditional subject thesaurus style searching. As a result, they may be even further deceived about the efficacy of free-text searching in the LEXIS and WESTLAW databases. The problem widens because any lawyer with her own personal computer and modem⁷⁹ can now dial in to either system to use the database. Because the personal computer will lack a "Deluxe" dedicated keyboard, using a PC would seem to force lawyers to confront their lack of computer literacy and warn them that their use of the systems may be inefficient. However, software developments are running apace, and soon it will be possible to buy a reasonably priced software package that allows the lawyer to interact with a "shell" program that is quite simple to use.⁸⁰ Thus, the same problems will occur on a wider scale.

The success of the moron Cadillac did not go unnoticed, and West eventually changed its strategy and marketed its own user-friendly terminal "WALT." (West Publishing Company held a nationwide contest to name the new user-friendly terminal. They sought a warm, avuncular name. See Woxland, *Anthropomorphism and the WESTLAW Custom Terminal OR "Hi Margie, This is Tom. It's About WALT . . ."*, LEGAL REFER. SERV. Q., Winter 1983, at 89.) Although still not as simple to operate as the LEXIS Deluxe terminal, the WALT terminal was a step towards easing the need for mechanical skills. Of course, some degree of compatibility with other information systems and low price was lost in this trade.

79. A modem is a device used for communication between computers over standard telephone lines.

80. Such software will allow the user to chose commands from "menu" screens that briefly explain the effect or result of each command. The software will then translate these commands so that LEXIS or WESTLAW understands them. Lacking such software, only the intrepid can figure out how to proceed.

Because the purveyors of the on-line databases generate income by charging per search and per unit of time that the system is in use, it is in their interest to encourage wide-spread terminal operation. It should come as no surprise that the marketing strategies of LEXIS and WEST-LAW have centered on every lawyer having his own terminal. This has only exacerbated the problem of inefficient searching by non-expert end-users.

3. *Some Theoretical Implications of the New Paradigm*

The full-text on-line legal databases are a new form of legal literature. The new literature is more or less identical in content to the old West system, but it is accessible in an entirely new way. If we concentrate on the notion of access to the case law, we can begin to understand how radically the legal databases break with the literature of the past.

The Digest was the internal, mediating structure within the old mode of discourse. The West editors were, in effect, the Platonic Guardians⁸¹ of legal language and legal meanings. The discourse, in turn, was the ground of integration and coherence in substantive law. The very notion that it was appropriate to place cases arising in state jurisdictions into a national index and national categories betrayed an underlying jurisprudence, a non-positivist view of the nature of law.

The location of issues and cases in the old paradigm was part of their meaning. Because the cases were only accessible through the Digest, they were always presented to the practitioner as situated. The situation was a substantive context, a setting that told the searcher the meaning of the case as much as did the opinion itself.

Free-text searching in legal databases, however, deprives the researcher of context. The materials are presented in a mechanical and (given the deficiencies of searching outlined above) an almost arbitrary fashion. Found cases that are relevant are like prizes in a computer game, rather than instantiations of the legally and socially appropriate categories of the West Digest. For example, in the legal databases the notion of making or deciding law by analogy is no longer a part of the primary source material itself, but must be added onto the raw data by the practitioner. Analogy has been a primary mode of legal discourse, and a primary instrumental technique for those advocating changes in the law.⁸² The Digest categories were themselves suggestive of analogues, but the simultaneous occurrence of search terms is not.

81. See L. HAND, *THE BILL OF RIGHTS* 73 (1958).

82. See Childress, *supra* note 8, at 1534 (arguing that on-line databases will discourage reasoning by analogy, and focus litigation practice on arguing with only "on point" cases, thus stifling growth and development in the law).

One way of thinking about the structural differences of the old and new paradigms is to think about legal research as a sort of economy. The goods exchanged in the legal research marketplace are the contents of the cases. I don't mean to suggest the literal research marketplace where information is available to everyone in proportion to the amount of money they have to pay for talented researchers, to pay for experts to weed out irrelevant retrieved information, and to pay for the necessary computer time, although that's certainly an important issue.⁸³ What I mean, rather, is an economy based on the "exchange" of information from the corpus of law (the sellers) to practitioners (the buyers).

The West Digest System was like a centrally planned economy. The practitioner could not obtain information directly from the cases, but was forced to go through the regulating mechanism of the Digest. This system was "efficient" because there were no alternatives; the buyer (practitioner) could not *find* the seller (sources of information) in the absence of the Digest. Also, the system was relatively leveling and egalitarian; it held fewer rewards for pure searching skill than does free-text searching. Reasonably competent searchers were able to find most relevant information, and only somewhat less relevant information than a very good searcher.

The West and LEXIS computer systems substitute a kind of marketplace for the planned economy of the Digest. The practitioner can obtain information directly from the cases by means of Boolean search techniques without reference to a central authority.⁸⁴ The overall efficiency of the system is questionable, although there are certain kinds of exchanges that it facilitates far better than the old system (e.g., finding all cases referring to a certain statute), and there are other kinds of exchanges that were impossible in the old system but now are quite simple (e.g., finding cases by judicial author).

There are several implications of this new form of exchange. First, the researcher with more skill can obtain a lot more information than

83. See Childress, *supra* note 8, at 1532 (pointing out that the financial costs of computer research make it available only to wealthy participants in the legal system, and arguing that this research advantage, e.g., more up-to-date Shepardizing, can raise ethical problems).

84. The full-text databases are subversive of authority in a much more direct and less metaphorical sense. The California Supreme Court "depublishes" opinions of lower courts that it disapproves without actually overruling the case or vacating the judgment. See *supra* note 27. Depublished opinions are not included in California's official reporters, and West does not insert headnotes from depublished cases in the American Digest System. They may not be cited and have no precedential value — they exist only in a kind of legal limbo. But Mead Data does not remove these opinions from their database. And, for some reason, enough lawyers have clamored for access to these cases that West has put them on WESTLAW!

those less skillful. Thus, the new system differentiates researchers based on merit; it rewards skill far more than did the old system. Second, the new system encourages "legal realist" practice, because it enables the practitioner to acquire and analyze cases by judge, by opposing counsel and by opposing party. Third, the new system can result in pluralistic legal discourse. The old system almost guaranteed that opposing counsel would be using the same source materials. The inefficiency of the new system in finding the relevant cases available through the West Digest⁸⁵ and its ability to pull in arguably relevant cases from anywhere in the corpus could result in different source materials for opposing counsel.⁸⁶ Counsel may end up talking past instead of arguing against each other, and a judge may be forced to choose the cases she prefers rather than the arguments she prefers.

Of course, interpretation of the meaning of cases, assessment of relevance, and analogy from rules and facts should result primarily from the professional abilities of the practitioner rather than the structure of the legal sources. This is true whether the raw materials of research (the cases) are obtained from the Digest or from a database. Also, the prudent practitioner will use both research sources, so that many if not most important cases will be discovered in the context provided by the old paradigm.⁸⁷ Still, the new paradigm is bound to influence the practice of law.

III. CONCLUSION

Does all this mean that computers should have no place in the research process? The answer quite clearly is no. To use on-line searching efficiently in the short term, lawyers have to develop strategies for dealing with its limitations. In the long term, the old and new paradigms will merge in the technologies of the future.

85. See *supra* text accompanying notes 48-72.

86. But see Childress, *supra* note 8. Childress suggests that use of the computer databases will tend to narrow the focus of practitioners to the "on point" cases rather than expand it in unpredictable ways, as I have suggested. *Id.* at 1534. This seems to be based on his belief that computer searchers will only find cases with matched facts and/or matched holdings. My guess is that, due to the shortcomings of free-text searching outlined above, even "on point" cases would sometimes not be retrieved. On the other hand, much that would be retrieved would appear relevant to the practitioner because it was *retrieved*, rather than because it was relevant. The result would be pluralistic discourse.

87. See *supra* note 47.

A. The Short Term: the Use of Computers as an Adjunct to Traditional Research

The strengths and shortcomings of the old and the new forms of legal literature are complementary. Prudent lawyers will continue to use both manual hard copy research and on-line free-text searching side-by-side. As I have suggested, the special difficulties and limitations of computer research mean the average lawyer is not the optimal end-user of the system.

Because of law school training programs, law firms increasingly will be composed of lawyers who know that LEXIS and WESTLAW are powerful tools. But lawyers will not maintain and expand the search skills necessary to use the databases efficiently. Attorneys will do computer research as they have done hard copy research — senior members of firms will refer problems to the newest lawyers on the staff who, because of their recent graduation from law school, will be more familiar with the computer systems. In my view, however, this practice is an inadequate response to the need for special skills and constant updating.

Law firms have to recognize that the average attorney has no time to maintain his or her on-line research skills. The attorney who is a true computer "jock" is an exception. As the marketing struggle between LEXIS and WESTLAW continues, they will offer more databases with ever-expanding search capabilities. As the systems become more complex, lawyers will need to continually develop more sophisticated computer skills. Law firms will reach a point where they must decide to create a new professional position: an expert in computer research.

Attorneys must recognize the need for an expert who straddles the law librarian's function and associate/researcher's function. This new position may require individuals trained both in law or legal research and in librarianship and computer technology. These intermediaries must be able to fully understand a lawyer as he or she describes a problem, and then be able to employ their in-depth understanding of the databases to formulate effective searches and to retrieve relevant information. Some large law firms are beginning to recognize that computer-literate law librarians are well-suited to fill the role of the LEXIS/WESTLAW expert.

The Berkeley law library can serve as a model for this kind of system. One member of our reference staff, who is both a lawyer and a librarian, is responsible for maintaining current knowledge of the databases. Even those of us who think of ourselves as computer literate, and who use the databases with some frequency, cannot hope to keep up with day to day changes and evolutions in the systems. Instead, we rely on this individual to keep us posted on the steady stream of updates. Law firms will have to recognize the need for this type of specialist and

to compensate them accordingly. Until then, the enormous potential of the on-line systems will be diluted or lost, and standard of care questions centering on ineffective and inept use of the systems may arise in the near future.

B. The Long Term: Replacing Traditional Research with Enhanced Computerized Research

The perceptive reader will have noticed that all of my criticisms of computer searching hinge on the use of free-text searching. This is a significant limitation. Computers can accomplish many traditional research tasks more efficiently than hard copy products. For example, the on-line version of Shepard's Citations is always much more up to date than the shelf versions, it contains all the relevant citations in one place (rather than across several volumes), and the on-line version allows the researcher to jump quickly back and forth between the citator and the cases.⁸⁸ Since all of the references in Shepard's need to be checked (although most are irrelevant), the on-line researcher does not need to spend most of her time running around the library and physically locating the cases. The on-line citations systems being implemented by Mead Data (Auto-Cite) and West (Insta-Cite) are powerful new tools in the research world as well.

One quick and dirty way of improving the computer systems would be to load the entire West Digest System on-line.⁸⁹ I have already acknowledged the importance of the ability of the on-line systems to search for unique words or terms with free-text searching. But for efficient research in the realm of legal concepts, the intervention of a highly trained and dependable human indexer may be irreplaceable. The Digest could be used both as a subject for free-text searching and for browsing like a traditional subject thesaurus. All the cases could be cross-referenced with the index so that the researcher scanning the all-inclusive Digest (no more searching from decennial to decennial) could jump immediately to the indexed cases and back, much like the current Shepard's citator. The experience, vocabulary and expertise of the West editors would be an extremely useful addition to on-line research.

Over time, computers may begin to replace hard copy as the medium for traditional index-based research through the convergence of emerging technologies. For example, consider the recent advent of CD-

88. See Dabney, *supra* note 24, at 38-39.

89. The West "Full-Text Plus" system is supposed to provide many of these benefits, but it provides them only in the context of free-text searching, and there are other significant limitations. See *supra* note 68.

ROMs.⁹⁰ CD-ROM technology allows small computers to use the laser disks used by home stereo compact disk players as optical storage devices. A single CD-ROM will hold 600 megabytes of information⁹¹ (about 300,000 typescript pages), the equivalent of about 150 volumes of *United States Reports*. The last fifty years of the entire West National Reporter System would fit on fifty or sixty CDs, which could be easily stored in a small desk drawer. Never before has so much information been available in such a small and physically stable format. Also, manufacturing costs of CD-ROMs are low; each disk costs less than ten dollars to produce in quantity.⁹² Although many users don't realize it, LegalTrac (a computer-based legal periodical index marketed by Information Access Company) is a CD-ROM device that has already arrived in many law libraries. In the law office of the future, the firm computer will provide reporters and digests on-line, and lawyers will be able to use both powerful indexes and free-text searching to access the database.⁹³

90. Compact Disk - Read Only Memory. A CD-ROM is a small (4.72-inch-diameter) plastic-coated metal disk with binary information etched onto the metal. The information is read by a "player" disk drive that bounces a small laser beam off the surface of the disk and reads the modulations. Information can be read from the disk but not written to it.

91. Crabb, *CD-ROM Arrives; It's Fast but Limited*, InfoWorld, Mar. 31, 1986, at 49, col. 1 (evaluating performance of Phillips CD-ROM device and the few software packages currently available).

92. Compare these manufacturing costs to the retail prices of the hard copy products, *supra* note 43, and you will see that replacing hard-copy products with CD-ROMs could mean both enormous profit margins for West and significantly lower prices to legal consumers. The CD-ROM disk drives are also quite inexpensive. Single disk drives now cost as little as \$845, and should soon be available for under \$500. Welch, *Manufacturers to Propose CD-ROM File Standard*, InfoWorld, Feb. 3, 1986, at 1, col. 1. For more on CD-ROMs and the attempt to develop industry standards in order to facilitate market expansion, see *id.*

93. Of course the paperless office is a myth. People will always want to interact with words on paper. But a computer-based library is not incompatible with the desire for printed products. First of all, the quality of computer visuals is skyrocketing. Personal computers and terminals of the near future will have very large and very high-resolution screens. Legal materials will appear on-screen in a life-size black on white replication of actual typeset pages. The user will be able to instantaneously "thumb" through pages as easily as through a book. This will make scanning search results far faster, easier, and more natural than current dumb terminal technology. Second, law offices will see the emergence of a technology called "on-demand publishing." High-speed laser printers in the library will typeset cases and other materials onto book-quality paper in a matter of seconds. The paper output can be bound or inserted into looseleaf binders.

Thus, the entire contents of a legal library will be on-line, with most of the collection on a local computer (e.g., on CD-ROMs updated every month or two), and very recent materials available from remote sources such as LEXIS and WESTLAW. Portions of the collection that are used frequently will be "published" in hard copy and bound by the library staff. The printed collection can be instantly updated in order to make information access easier and to give legal workers the printed page that they will long desire.

For now, LEXIS and WESTLAW will remain part of the armory of legal research — useful, practical, and particularly helpful for questions that call for the location of a specific word or name. And we should not underestimate the role of computers and free-text searching in the law library of the future. But unless and until subject thesauri implemented by professional indexers are added to the databases, these systems will not be ultimate research tools. Until that time, lawyers must be careful that the shortcomings of the on-line legal literature do not distort or diminish the quality of their practice.