

HEINONLINE

Citation: 78 Law. Libr. J. 5 1986

Provided by:

Maurer School of Law Library



Content downloaded/printed from

HeinOnline (<http://heinonline.org>)

Wed Mar 9 13:05:11 2016

-- Your use of this HeinOnline PDF indicates your acceptance
of HeinOnline's Terms and Conditions of the license
agreement available at <http://heinonline.org/HOL/License>

-- The search text of this PDF is generated from
uncorrected OCR text.

-- To obtain permission to use this article beyond the scope
of your HeinOnline license, please use:

[https://www.copyright.com/ccc/basicSearch.do?
&operation=go&searchType=0
&lastSearch=simple&all=on&titleOrStdNo=0023-9283](https://www.copyright.com/ccc/basicSearch.do?&operation=go&searchType=0&lastSearch=simple&all=on&titleOrStdNo=0023-9283)

The Curse of Thamus: An Analysis of Full-Text Legal Document Retrieval*

Daniel P. Dabney**

Full-text computer-assisted research systems have become standard tools for searching large quantities of legal documents. Yet there remain questions as to the relative effectiveness of full-text searching. Mr. Dabney reviews recent research into these questions and discusses the implications of the results for computer-assisted legal research systems. He concludes that the performance of currently available systems could be improved.

Contents

I.	Introduction	5
II.	A Basic Model for Document Retrieval Systems	7
III.	Basic Concepts of Conventional Indexing	9
IV.	Key Ideas in the Evaluation of Document Retrieval Systems	14
V.	Introduction to Full-Text Document Retrieval Systems.....	17
VI.	File Size and Retrieval Performance in Full-Text Systems ...	21
VII.	The Experiment of Blair and Maron	26
VIII.	WESTLAW and "Full-Text Plus"	31
IX.	Ramifications for the Users of CALR Systems	34
X.	Conclusion	38

I. Introduction

Plato gives us the legend of the Egyptian god Theuth, giver of marvelous inventions. Theuth invented geometry and astronomy, games and dice, but his greatest invention was writing. The king of the Egyptians, Thamus, admired many of the gifts of Theuth, but he did not approve of writing, and refused to teach the art to his subjects.

* © Daniel P. Dabney, 1986. This is an edited version of a paper presented at the 78th Annual Meeting of the American Association of Law Libraries, New York, New York, July 9, 1985. It is one of the winning articles in the 1985 Call for Papers competition addressed to newer law librarians.

** Reference Librarian, University of Texas Tarlton Law Library, Austin, Texas. The author would like to thank M.E. Maron and David Blair for their kindness in supplying an advance copy of their article and supplemental material, and Robert C. Berring and Roy M. Mersky for their support and encouragement. *Editor's note:* Responses to this article from representatives of Mead Data Central and West Publishing Company will be included in the next issue of *Law Library Journal*.

[‘]If men learn this, it will implant forgetfulness in their souls; they will cease to exercise memory because they rely on that which is written, calling things to remembrance no longer from within themselves, but by means of external marks. What you have discovered is a recipe not for memory, but for reminder. And it is no true wisdom that you offer your disciples, but only its semblance, for by telling them of many things without teaching them you will make them seem to know much, while for the most part they know nothing, and as men filled, not with wisdom, but with the conceit of wisdom, they will be a burden to their fellows.’¹

Our legal system has produced millions of writings concerning the law, and thousands of new writings are created each day. How do we deal with the objection of Thamus? Since the mere possession of writings does not give knowledge, how are we to extract from this almost incomprehensibly large collection of written records the knowledge that we need?

Historically, document retrieval systems have relied upon the skills of indexers. We cannot read all of the documents that might contain relevant information, so we rely on others to read the documents for us, and to note for us the texts that we will need to consult in the future. But human indexing is fraught with error and uncertainty, so we have sought other ways to extract information from written texts.

For the last twenty-five years, there has been increasing interest in and reliance on an alternative approach to document retrieval. Instead of relying on the notations of human indexers, users of full-text information systems rely upon the power of computers to find desired documents by examining their unindexed texts. Legal researchers have been among the first and heaviest users of such systems.

But questions remain about the effectiveness of full-text computer-assisted legal research (CALR) systems, and new information is becoming available. In a recently published paper, David Blair and M. E. Maron reported the results of an experiment that may have far-reaching implications for those who do legal research.² In this first experimental test of a large-scale, full-text document retrieval system, Blair and Maron found that the system retrieved no more than 20 percent of the total number of documents relevant to sample search queries. The time has come for a reassessment of full-text document retrieval as a legal research tool. This article is an attempt at this reassessment.

One of the primary purposes of this article is to draw the attention of the legal community to the findings of the important experiment performed by Blair and Maron. Beyond this, the paper does not present any original research on the effectiveness of document retrieval systems. None of the author's small experiments mentioned in this article are intended to do more

1. PLATO, PHAEDRUS 275a-b.

2. Blair & Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System*, 28 COM. A.C.M. 289 (1985) (publication of the Association for Computing Machinery).

than illustrate a qualitative feature of CALR systems. There is clearly a need for more information about the performance of existing systems. Such a study is no more than suggested here, however; the reader should not attempt to draw quantitative conclusions from this article.

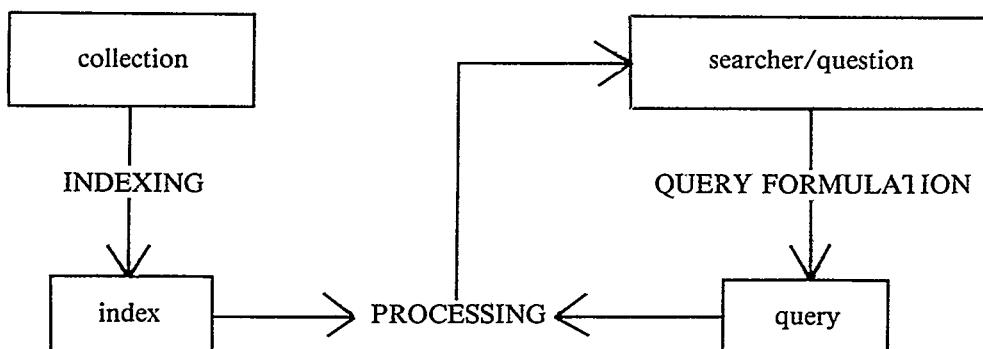
Although this article is not intended to suggest improvements to existing CALR systems, it is impossible to critique the existing systems without implying that changes are needed. It is hoped that the discussion here will be of use to those who consider such changes. But establishing new systems or improving old ones raises many questions beyond what is discussed here.

This article deals only with document retrieval systems. Information comes in many forms other than retrieved documents, and many of the most interesting applications of information science to the practice of law involve information in these other forms. Even within the limited sphere of document retrieval, this article is focused on the full-text retrieval technique and discusses other methods only for the purpose of highlighting the distinctive features of full-text systems. The retrieval effectiveness of other systems is not discussed, so no conclusions about the comparative value of such systems are to be drawn.

Finally, although the paper attempts to assess the impact of full-text retrieval systems on the practice of legal research, it does not attempt to assess CALR's impact on the practice of law generally or on the administration of justice.

II. A Basic Model for Document Retrieval Systems

Few people take the position argued by Thamus, that writing should be avoided because it corrupts rather than enhances memory; most now agree that writing has some part to play in document retrieval. But beyond this there is little agreement. What kinds of writing are most useful for insuring that needed documents will be findable in the future? To examine the similarities and differences of the various techniques now employed, it is helpful to consider a simple model of information systems generally, which may be illustrated thus:



At the outset, any document retrieval system needs a group of documents, the collection, and a need for information from those documents, a searcher with a question. In a system with a small number of documents, nothing else is necessary. For each question, the searcher can examine each document in the collection and select the relevant ones without relying on any recognizable system. But where the size of the collection prohibits this simple approach, the process of document retrieval is broken down into three suboperations: indexing, query formulation, and processing.

Indexing may be broadly defined as the processing of the documents in the collection that takes place before the searcher approaches the system with a question. It includes whatever must be done with the documents to make them findable by the other operations of the system. Whatever results from this process is the index. For the remainder of this paper, however, indexing will be used to refer to conventional indexing. A typical conventional subject index is a list of subject headings with references to the documents in which the subjects are discussed, and conventional indexing is the process of examining documents for the purpose of making such postings. Conventional indexing is discussed at greater length in the next section.

Query formulation is the portion of the retrieval system that is left to the user. The user of a conventional subject index is performing query formulation in selecting subject headings to consult.

Processing is the operation in which the query is matched against the index according to whatever retrieval rule is in use in the system. Processing in a conventional document retrieval system involves the searcher's turning pages to look for subject headings corresponding to the query.

Considered in terms of this model, the differences between full-text CALR systems and conventional subject indexes are striking. In conventional systems, the query formulation and processing operations are relatively simple—so simple that it is easy to overlook them when considering the retrieval system as a whole. Conventional systems stand or fall according to the quality of the indexing that goes into them. In contrast, full-text systems contain little in the way of indexing, but depend heavily on quality of the query formulation provided by the user and on the power of their processing algorithms.

One of the chief factors that motivates the use of full-text retrieval systems is a distrust of the indexing process. Conventional subject indexing is a difficult and error-prone enterprise, and any document retrieval system that is able to dispense with such indexing avoids many kinds of problems. To understand the relative merit of full-text systems, it is first necessary to consider some basic concepts of the indexing it attempts to replace. What follows is a short description of some key features of indexing, with references to some conventional indexing systems familiar to legal researchers.

III. Basic Concepts of Conventional Indexing

Conventional indexing typically involves having an indexer examine each document as it is added to the collection and enter references to that document in an index. The indexer identifies the salient features of a document and posts references to those features in a conveniently arranged list of possible features.

One of the most important features of a document is what it is about—this is subject indexing. Subject indexing is one of the most useful forms of indexing, but it is also the most difficult. Although indexing documents by other salient features, such as author and title, is subject to much more difficulty than the uninitiated would imagine, the following discussion deals primarily with the issues that arise in subject indexing.

A. Document-Oriented Indexing and Request-Oriented Indexing

The central questions for an indexer to answer are which properties of the document being indexed are the salient ones, and how should these characteristics be represented in the index? According to the more prevalent practice, the job of the indexer is to capture the essence of the document being indexed. The basic tenet of document-oriented indexing is that the salient characteristics of a document are intrinsic to the document. This approach has a certain aesthetic appeal, but most indexing theorists see it as misguided.³ Indexing is not done for its own sake, but to facilitate a document retrieval system. Thus, the salient characteristics of a document are not intrinsic to the document being indexed, but are relative to the needs of the users of the system. This is the basic concept of request-oriented indexing.

To illustrate the difference between document-oriented indexing and request-oriented indexing, consider the following example. A case is decided that deals with the question of whether the doctrine of *cy pres* may be used to reform a grant made in a will so as to avoid the invalidating effect of the rule against perpetuities. The document-oriented indexer goes to the essence of the case, probably choosing terms like “perpetuities, rule against” and “*cy pres*.” The request-oriented indexer takes a different approach, asking who is likely to want to see this case. Because an attorney who wants to see that case may not have thought of using the doctrine of *cy pres* as an escape device, the indexer uses the heading “perpetuities, rule against—escape devices” or “perpetuities, rule against—ways to beat it.” From a different point of view, the indexer might anticipate that a law professor would be glad to find the case under “*cy pres*—peculiar applications.”

3. See, e.g., D. SOERGEL, INDEXING LANGUAGES AND THESAURI: CONSTRUCTION AND MAINTENANCE, 45-50 (1974).

Most of the indexing schemes in use today are clearly document-oriented. Consider, for example, Library of Congress subject cataloging: "The aim of subject cataloging at the Library of Congress is to summarize and describe the content of a work by the assignment of subject headings."⁴ Similarly, the West key number system is based upon the assignment of key numbers according to the content of the opinion rather than its usefulness for anticipated questions.⁵

It may seem peculiar that request-oriented indexing, which enjoys such a distinct theoretical advantage over document-oriented indexing, is so little practiced. The reasons for this are partly historical and partly practical. Historically, request-oriented indexing was virtually unknown in the late nineteenth century, when the great indexing systems now generally used were being developed. The revision of a general indexing system is such a titanic project that it is seldom undertaken, and the systems of the past tend to be propagated into the future. Other reasons for the continuance of existing indexing systems are that the development of them already has been paid for and that they are familiar. For example, in the creation of a new index of legal periodical literature (the *Current Law Index* and the *Legal Resource Index*), the American Association of Law Libraries chose to adopt a slightly augmented form of the Library of Congress subject headings. Adopting the indexing technology of the turn of the century for a new on-line system may seem extremely foolish, but the Library of Congress subject headings were both familiar and available.

The practical reason for not adopting request-oriented indexing is that it is more difficult to do—in many ways, impossibly difficult. The indexer has limited knowledge as to what features of a document are going to be useful to later researchers. To the extent that the indexer does not properly anticipate the use of the system, the indexing will be ineffective. Even if the indexer knows all possible questions that might be submitted to a retrieval system, the process of testing each document against every one of those questions is unworkably time-consuming.

The concept of request-oriented indexing, however, is compelling enough that some document-oriented systems have request-oriented features. For example, Anglo-American cataloging rule 21.29B directs the cataloger to put an added entry for a document under any heading that the indexer supposes might be used by a searcher who is looking for the document being indexed.⁶ Here, the cataloger is being asked to anticipate the questions of future searchers and to index accordingly.

4. P. ENYINGI, M. LEMBKE & R. MITTAN, CATALOGING LEGAL LITERATURE: A MANUAL ON AACR2 AND LIBRARY OF CONGRESS SUBJECT HEADINGS FOR LEGAL MATERIALS 329 (1984).

5. See J. JACOBSTEIN & R. MERSKY, FUNDAMENTALS OF LEGAL RESEARCH 66 (3d ed. 1985).

6. AMERICAN LIBRARY ASSOCIATION, ANGLO-AMERICAN CATALOGING RULES 322 (2d ed. 1978).

B. Whole Document Indexing and Analytical Indexing

Another key decision to be made in indexing is the size of the unit to be indexed. Whole document indexing considers each document as a unit, to be represented in the index by entries covering it in its entirety (if possible). Whole document indexing is almost always document-oriented.

The outstanding example of whole document indexing is Library of Congress classification. Since a document can only be shelved in one place in a library, it can have only one call number assigned to it, and that call number has to represent the entire document as best it can.

Library of Congress subject cataloging also is based to a considerable extent on whole documents. Most books are represented in subject catalogs by a single subject heading that is (ideally) coextensive with the content of the book. Where this is not possible, it is permissible to use two or three subject headings, but it is considered bad practice to use more than three.⁷

Analytical indexing is based on a unit that can be somewhat smaller than a document. Request-oriented indexing tends to be analytical because different parts of a document may have different anticipated uses, but analytical indexing also may be nothing more than document-oriented indexing done on a finer level.

Analytical indexing has obvious virtues. An index that contains analytical entries contains more information than one that is tied to document-long units. The West key number system is analytical because it takes as its unit "points of law" rather than entire cases. Since cases commonly discuss several otherwise unrelated points of law, analytical indexing is clearly necessary.

C. The Role of Authority in Indexing

Most indexing systems make use of some form of authority control for the purpose of bringing together similar postings in the index. Index users expect to find all of the documents that share a salient characteristic indexed in the same way. For example, legal researchers may not know whether they will find documents relating to children under "children," "minors," "juveniles," or something else, but they do expect that all of the entries will be in one place.

Without making light of the difficulties encountered in work with name authority, uniform titles, and other forms of authority control, it may be said that the most difficult form of authority to establish is subject authority. The authority control list for a subject index is called a thesaurus, and it is the thesaurus that determines the universe of possible indexing entries for a system.⁸

7. P. ENYINGI, M. LEMBKE & R. MITTAN, *supra* note 4, at 358-59.

8. Another effect of subject authority control in indexing may be an influence on the substantive

In a document-oriented indexing scheme, the thesaurus is the universe of possible topics for documents; in a request-oriented indexing system, it is the universe of possible questions. The effectiveness of an indexing system, then, is limited by the quality of its thesaurus, and this can be a severe limitation indeed.⁹

D. Precoordinated and Postcoordinated Indexes

One of the most telling features of an index is the way it deals with subject entries based upon more than one idea. Consider again the problem of indexing a case that discusses the use of the doctrine of *cy pres* to escape the operation of the rule against perpetuities. Both “*cy pres*” and “perpetuities” are obvious choices for index entries, but neither entry by itself conveys enough information to give an adequate representation of the document in an index entry.

In printed indexes, the terms need to be stacked, or (in the jargon of indexing theory) “precoordinated.” The index entry for such an item will be either “perpetuities—*cy pres*” or “*cy pres*—perpetuities.” Whichever form is chosen, some users of the index are bound to look under the wrong term, and thus have difficulty finding the germane entry.

This problem can be avoided to some extent by posting a reference to the case under both terms, though this increases the size of the index. If three terms are stacked in the index entry, it would be necessary to make six postings in the index to cover each different stacking of the terms; for four terms, twenty-four postings would be required, and so on. Printed indexes are usually limited in their ability to offer different stackings of index terms.

development of the subject of the collection. For example, some of the terms that might be used as subject headings have connotations that implicitly comment on the subject matter so indexed. Consider, for example, that generations of lawyers and judges have found law relating to employment relations under the heading “Master and Servant.” This subject heading no doubt seemed reasonable to the legal community of the turn of the century when the heading was incorporated into the West key number system. A different segment of the society of that period might have found it reasonable to put such material under the heading “Toiler and Leech,” and colored future perception of the topic in a different way. “Toiler and Leech” seems outrageous to us; “Master and Servant” seems merely archaic, but this is to a large extent the effect of familiarity. For an artful demonstration of differing levels of perceived prejudice in language, see D. HOFSTADTER, *A Person Paper on Purity in Language*, in METAMAGICAL THEMAS 159 (1985). Many indexing systems make some effort to eliminate bias in their subject headings. For example, between the Eighth and Ninth Decennial Digests, the topic “Bastards” was redesignated “Illegitimate Persons” and has since been changed to “Children-out-of-Wedlock.” The precoordination of subject headings in a thesaurus also may affect the development of the literature by making it appear that certain ideas go together and others do not.

9. The greatest objection to the older version of the *Index to Legal Periodicals* was that it was based on an inadequate thesaurus, one that contained too few subject headings to represent the topics covered by its collection. See, e.g., *Report of the Subcommittee on the Index to Legal Periodicals*, in PROCEEDINGS OF THE 1976 ANNUAL MEETING OF THE ASSOCIATION OF AMERICAN LAW SCHOOLS, pt. 1, § 1, at 30, 33-34.

The problems encountered in combining index terms are illustrated by the structure of the West digests, the primary indexes for American case law. For example, one case¹⁰ makes the point that the burden of proving the existence of an exemption from Arizona's blue sky laws rests upon the party raising the defense. This is not a concept that can be attached nicely to a single index term, and it is not surprising that West's primary indexing, the assignment of a key number, involves a considerable stack of subject headings:

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

This full stack is represented by West key number Securities Regulation 327. The complexity of legal documents requires an indexing scheme that is capable of reflecting several ideas at once, that is, a "deep" index. Deep stacking is a feature of the West key number system throughout.

West's key number system is admirable in many ways, but there are clearly certain problems with it. The five-tiered indexing of this case, for example, does not include references to either "burden of proof" or "exemptions," two ideas that are essential to the point of law being discussed. The digest system cannot possibly represent all of the salient features of a case in its indexing, for this would require there to be almost as many key numbers as there are cases. This is a limitation on the possible depth of indexing.

The second difficulty is that even this limited stacking of index terms can be difficult to manage. West prints each headnote in only one place in the digest (or, in rare instances, two places). The user who wants to find the case mentioned above must discover that the first indexing element used was "securities regulation," rather than, for example, "criminal law" or "evidence." The searcher must correctly anticipate that the next indexing element was "state regulation (blue sky laws)," and so forth, down to the bottom of the stack.

To facilitate this sort of searching, West prints an outline of the key number system at the head of every topic in the digest, together with scope notes that refer the reader to related topics. In practice, the searcher ordinarily needs only to guess the first two or three headings used by the key number system and can find the finer subdivisions by browsing.

Another feature that facilitates the use of the digest is the presence of a secondary index—in effect, an index to the index. The "descriptive word index" (DWI) refers the reader to appropriate key numbers from terms that

10. *State v. Baumann*, 125 Ariz. 404, 610 P.2d 38 (1980).

are not the chosen entries for the key number system itself. Terms in the DWI are necessarily less deeply stacked, however, and this approach is not always successful.¹¹

All of these problems can be traced to the difficulties of precoordinate indexing. The thesaurus of the indexing system has to preordain the manner in which multiple ideas will be combined into single index postings. It is necessary to stack the index terms because the index itself is a "linear file" in which each combined heading needs to have a unique location determined by the ordering of its parts.

At least in theory, a great many of these problems disappear if it is not necessary to post the index entries to a linear file. If the index entry need not be printed in any one place determined by its index terms, the indexer may index as deeply as desired (by assigning as many index terms as desired). This is called "postcoordinate indexing." True postcoordinate indexing requires the document retrieval system to have some means of retrieving combinations of these separately assigned subject headings.¹² This capability is ordinarily provided by a computer.

This short review of ideas in indexing shows that the indexing process is prone to many sorts of errors and uncertainties. Manual indexing is only as good as the ability of the indexer to anticipate questions to which the indexed document might be found relevant. It is limited by the quality of its thesaurus. It is necessarily precoordinated and is thus also limited in its depth. Finally, like any human enterprise, it is not always done as well as it might be.

Full-text document retrieval was devised to avoid these problems.

Traditional legal research procedures are rapidly proving inadequate to permit access to vast, continually expanding reservoirs of information. Based largely in the hierarchical organization of subject matter, manual research tools are effective only so long as the lawyer can easily tune in on the mental frequency of the person who indexed the information the lawyer seeks. While this system has previously been sufficient to meet most of lawyers' research needs, it has grown too cumbersome, too expensive and too rigid to accommodate practically and efficiently either the continuous influx of routine material or such new precedent as lawyers and judges are now formulating in evolving areas of law.¹³

IV. Key Ideas in the Evaluation of Document Retrieval Systems

The remainder of this paper will examine the effectiveness of the proposed cure—document retrieval systems. To facilitate this discussion, this

11. In the case used for the preceding illustrations, for example, the DWI is unhelpful. Securities Regulation key number 327 is not posted under "securities regulation," "exemptions," "evidence," "burden of proof," "registration," or "criminal law" in the DWI.

12. Some systems that do not have any mechanism for allowing the user to specify combinations of index terms in a query still claim to be postcoordinated if they permit the assignment of multiple index terms. Such systems might better be described as "uncoordinated."

13. Legal Research and the Computer (1975) (early promotional material from LEXIS).

section will define the measures by which document retrieval systems are judged. These measures are recall, precision, and fallout.¹⁴ Central to the definitions of all three of these measures is the notion of relevance.

A. Relevance

Relevance—the relationship between a question and a document that makes the document important to the person researching the question—is an extremely difficult concept in the law. Article IV of the Federal Rules of Evidence, "Relevancy and Its Limits," together with annotations to interpreting cases, occupies 248 pages of the *United States Code Annotated*.¹⁵ These pages are full of disagreements among lawyers on questions of relevance, and it is clear that any study of document retrieval needs a simpler way of determining relevance.

Some document retrieval experiments have relied upon subject experts to judge the relevance of documents to questions. In one important study,¹⁶ four attorneys evaluated cases on a four-step scale: "on point," "relevant," "related," and "irrelevant." The majority of cases were considered irrelevant by all four of the judges, but among those that were found at least related by one judge, all four judges agreed only 4.3 percent of the time. More than twice as often (11 percent of the time) the decisions of the judges spanned the entire range from on point to irrelevant. The judgment of experts seems to be an unsound way to determine relevance.

In better studies, including Blair and Maron's, all relevance decisions are left to the searcher—the relevant documents are the ones that the person doing the research says are relevant. Adopting this definition avoids the problem posed by the inconsistency of "expert" relevance judgments, but it makes a single person responsible for all relevance evaluations for any question in the study.

B. Recall

Recall is the percentage of the total number of relevant documents in a data base that are retrieved by the search being studied. For example, if a collection contains twenty documents that are relevant to a given question and a search in the collection produces seven of the twenty, the recall for that search is 35 percent. The key finding of Blair and Maron is that the full-text retrieval system they studied had an average recall of no more than 20 percent.¹⁷

14. For a general introduction to both the terminology and the substance of information retrieval, see A. FOSKETT, THE SUBJECT APPROACH TO INFORMATION (4th ed. 1982).

15. 28 U.S.C.A. 119-367 (1984).

16. Eldridge, *An Appraisal of a Case Law Retrieval Project*, in PROCEEDINGS OF THE COMPUTERS AND THE LAW CONFERENCE 1968 at 36, 41 (D. Johnston ed.).

17. Blair & Maron, *supra* note 2, at 293.

The adversary system of American law puts a high premium on exhaustiveness in case research and thus needs retrieval systems that maximize recall. Routine legal research is expected to turn up all of the cases that might be cited by the other side in a lawsuit. Lawyers tend to think that a few cases will not suffice to illuminate an issue, but expect instead to find every case that has considered it. When a point of law has been considered in so many cases that it is impossible to read them all, lawyers still expect to find all of the cases from the home jurisdiction or all of the most recent cases, not just a representative sampling of the many cases available.

Recall is notoriously difficult to measure. The total number of relevant documents found by a search usually can be determined, but the number of relevant documents in a collection *not* found by a search is seldom known.

C. Precision

Precision is the percentage of the total number of retrieved documents that are relevant to the search question. For example, if a search retrieves ten documents, of which seven are relevant, precision is 70 percent.

Precision is not as important as recall in legal research, but it is still an important measure. Lawyers' time is expensive, and time spent reading irrelevant cases can be an economic hardship for clients. Unlike recall, precision is fairly easy to measure. Calculating precision does not require any information about the part of a collection not retrieved in a search.

One of the best-known principles of document retrieval is that there is a roughly inverse relationship between precision and recall. A narrowly drawn search that retrieves few cases ordinarily can be made to have relatively high precision, but a search that is made broader for the purpose of retrieving more relevant cases ordinarily has lower precision. Since lawyers value recall highly, one might expect that they would be willing to accept relatively low precision.¹⁸

D. Fallout

Fallout is the proportion of the total number of irrelevant documents in a collection retrieved by a search. Unlike recall and precision, what constitutes an acceptable value for fallout depends on the size of the collection. One percent fallout in a collection with 400 irrelevant documents would give the searcher only four irrelevant documents to read, but one percent fallout

18. Some experts recommend that users of CALR systems use searches broad enough to achieve about 50 percent precision to achieve a satisfactory level of recall. Sprowl, *WESTLAW vs LEXIS: Computer-Assisted Legal Research Comes of Age*, 15 PROGRAM 132, 135 (1981).

in a collection with 400,000 irrelevant documents would flood the searcher with 4,000 irrelevant documents. Because legal research data bases tend to be large, only extremely low levels of fallout are acceptable in these systems.

Like precision, fallout is relatively easy to calculate for individual searches. The number of irrelevant documents retrieved is usually known. The number of relevant documents in a collection is a trivial fraction of the entire collection, and so the total number of irrelevant documents is, for all practical purposes, about the same as the size of the collection, and this number is usually known.

V. Introduction to Full-Text Document Retrieval Systems

In full-text document retrieval, there is no human subject indexing.¹⁹ The usefulness of such systems depends upon the extent to which they provide an adequate substitute for human indexing. Although full-text systems are not subject to many of the shortcomings of conventional subject indexing, they are subject to several distinctive problems, which will be discussed presently.

Imagine a legal research system that operates this way. An attorney calls a law clerk into his or her office and explains the research problem, telling the clerk what sort of case authority is needed. The law clerk then reads the entire National Reporter System, from the first volume of the first series of each reporter through all of the most recent advance sheets. The clerk makes copies of all of the cases that seem to be relevant, and returns the stack of copies to the requesting attorney. In this wonderful system, there would be no errors caused by indexing because there would be no need for indexing—each document would be judged on its individual content rather than by its representation in indexing terms.

Human beings are much too slow in their reading of cases to make such a system practical, but it is possible to accomplish more or less the same result by using computers.²⁰ Full-text document retrieval systems realize the

19. This is not to say that there is no indexing of any kind in a full-text data base. Documents added to a full-text system are posted to an inverted file (a "concordance") that serves as the index of the system and greatly facilitates its operation. In addition to the file inversion, which is mechanical, most full-text systems identify the inverted file postings as being from a particular part of the document. For example, data bases containing case law typically have "field" or "segment" indicators for the name of the case, the name of the authoring judge, the date of the decision, and so forth. This is not an entirely mechanical process and so can be considered a use of human indexing. However it is viewed, it is one of the most useful features of the system.

20. Even the fastest computers currently available cannot make a linear scan of a body of text as large as the National Reporter System. Full-text search requests are processed using an inverted file of the words appearing in the collection. This distinction is ordinarily invisible to the user of the system, but it does much to explain why the systems are designed as they are.

virtues of such a system, but only to the extent that the computer can satisfactorily "read" the searched text, and separate the relevant documents from the irrelevant ones.

Much work has been done in the field of artificial intelligence on the problem of machine understanding of texts. While some progress has been made in this area, no close approximation to the human understanding of natural language has been achieved, and none is likely to come soon.²¹ The deficiencies of full-text document retrieval all stem from the machine's inability to read cases closely enough to distinguish relevant documents from irrelevant ones.²²

The full-text systems currently used for legal research are simple, employing little of what is known about machine understanding of natural language texts. Both LEXIS and WESTLAW rely almost exclusively on the ability of the systems to recognize words supplied by the user. The difficulty with this approach is that there is an imperfect correspondence between words and ideas. The specific problems that arise fall into three categories: synonymous words, ambiguous words, and complex expressions.²³

A. Synonymous Words

The first problem with word-matching in natural language texts is that in natural language many words can be used to refer to the same thing. For example, a court discussing a ten-year-old boy might refer to him by any one of a number of words, including:

boy	child	youth	infant
minor	juvenile	ten-year-old	young man

He also could be mentioned by a word indicating his relationship to the case itself or to some other person or institution such as:

son	brother	ward	student
pupil	victim	witness	plaintiff
defendant	appellant	petitioner	patient

Cases relevant to the same issue might refer to other people who have the same legal standing:

girl	daughter	ten-and-a-half-year-old
------	----------	-------------------------

21. For a review of research in this area, see Waltz, *The State of the Art in Natural-Language Understanding*, in *STRATEGIES FOR NATURAL LANGUAGE PROCESSING* (1982).

22. Some work has been done on document retrieval systems that avoid this problem by translating the information contained in the collection into a form that a computer can process, a knowledge representation language. See, e.g., C. HAFNER, *AN INFORMATION RETRIEVAL SYSTEM BASED ON A COMPUTER MODEL OF LEGAL KNOWLEDGE* (1981).

23. This taxonomy of errors, and some of the examples cited to illustrate it, are taken from J. JACOBSTEIN & R. MERSKY, *supra* note 5, at 435-37.

The court simply might refer to the boy by his proper name.

In formulating a search, it is necessary to avoid choosing search elements that can be expressed in too many ways. If the research question involving a ten-year-old boy has enough other features to limit a set of retrieved cases to a reasonable number, it may be possible to pick the cases involving boys out of the set by hand. If there are not enough other elements available, the question is unsuited to full-text research techniques.²⁴

It is simply not possible to anticipate all of the words that might be used to refer to a situation or idea in the text of an opinion. To the extent that the searcher does not correctly anticipate the language used to describe a matter of interest, relevant cases will be missed, and recall will suffer.²⁵

B. Ambiguous Words

Ambiguous words create the opposite problem. Just as one idea can be expressed in many different ways, one word can represent several different ideas. For example, a search for mentions of the drug DES (diethylstilbestrol) will retrieve cases that cite *Tinker v. Des Moines Independant Community School District*.²⁶ Problems of ambiguity cause poor performance in precision and fallout.

A serious problem with ambiguous words can make a search element unworkable. For example, a searcher looking for the word "court" used in the sense of "woo" is not likely to succeed in a full-text CALR search.

C. Complex Expressions

The most difficult problem for full-text systems is that much of the meaning of natural language texts is contained in complex relationships between words rather than in the individual words themselves. Simple word recognition of the kind done by LEXIS and WESTLAW cannot discern the meaning of complex expressions because the relationships between words in natural text are much more subtle than the connectors that may be used to join terms in CALR searches. Even the most sophisticated connectors in use for CALR relate words only by finding them in the same sentence or paragraph.

Consider, for example, this search 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 ques-

24. The selection of elements for a search is discussed further in section VI of this paper.

25. LEXIS and WESTLAW both provide limited assistance to the searcher by having the computer automatically search for words closely related to search terms (such as regular plurals). An expansion of this capability to include less obvious synonyms is seen by some as being a primary means for improving the performance of full-text systems. See Bing, *Third Generation Text Retrieval Systems*, 1 J. L. & INFORMATION SCI. 183, 191-93 (1983).

26. 393 U.S. 503 (1969).

tion, "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.

Where the words that express the elements of a search are sufficiently distinctive, the words themselves embody enough information to enable the computer to distinguish relevant cases from irrelevant ones. If, for example, one wants to find products liability cases involving the drug pitocin, the presence of the word "pitocin" in the case is a strong enough indication that the case will be of interest; the more complex syntactical relationships in the case's text are not needed for the search.

Certain kinds of questions tend to involve more distinctive words. Fact patterns, for example, tend to be relatively good for computer searching because they use words that do not appear in many cases (and because they are ordinarily less thoroughly indexed than are legal theories). Procedural questions, in contrast, are difficult to search because they tend to use words like "trial" and "motion" that occur often in legal literature.

Considered in the context of the discussion of conventional indexing systems discussed in section III, all of these problems may be traced to a lack of authority control. There is no controlled vocabulary for full-text searching. With respect to the other indexing concepts discussed in section III, the performance of full-text systems is mixed.

Full-text systems are, at least after a fashion, request-oriented. The selection logic of the search is almost entirely in the hands of the searcher, and the searcher can tune the search to his or her individual requirements. But making the searcher provide the selection logic also puts the success of the search more into the hands of the searcher than it is in conventional systems, and this may not be all good. A reputable publisher can make sure that all of the indexers it employs are skillful at their jobs, but not all users of research systems are experts.

Full-text systems are extremely analytical. When the searcher can find every significant word in the collection, no part of the collection is too small to be retrieved. This too, however, is a mixed blessing. Indexers identify the parts of a document that are most significant or important, but full-text systems uncritically treat all words as equals. Full-text systems make it possible to find bits of text that are beneath the notice of the indexers, but the more important features of the text may be obscured by the mass of detail.

Finally, full-text systems are wonderfully postcoordinated. There is practically no limit to the boolean magic that can be performed with them. But the use of conjunctive boolean connectors tends to hurt the retrieval performance of such systems, as will be seen in the next section.

VI. File Size and Retrieval Performance in Full-Text Systems

One of the factors that has a dramatic effect on the performance of full-text systems is the size of the collection being searched; the larger the collection, the poorer the performance. To assist in the discussion of this point, search results will be considered in terms of table 1.

Table 1

	relevant	not relevant	
retrieved	x	u	N_1
not retrieved	v	y	
	N_2		N

In this table, N is the total number of documents in the data base, N_1 is the number of documents retrieved in a search, and N_2 is the number of documents relevant to the search question. The N documents may be divided into four classes.

type x, relevant documents retrieved

type y, irrelevant documents not retrieved

type v, relevant documents not retrieved

type u, irrelevant documents retrieved

Recall for a search may be expressed as x/N_2 , the fraction of the total number of relevant documents that were retrieved by the search. Precision can similarly be expressed as x/N_1 , and fallout as $u/(N-N_2)$.

Retrieval systems make two sorts of errors, which may be characterized as *type u* and *type v*. Lawyers are particularly concerned with *type v* errors, relevant documents not retrieved, which contribute to poor recall. *Type u* errors, irrelevant documents retrieved, contribute to poor precision and fallout. *Type u* errors are not ordinarily thought to be of as much concern to attorneys, but they limit the quality of a search just as surely as *type v* errors.

Only a small number of *type u* errors in a system can be tolerated by a searcher. No one is ordinarily willing to read more than a few dozen irrelevant documents in response to a search. But *type u* errors can be eliminated only at the cost of creating additional *type v* errors. And since *type u* errors are related to the size of the collection, the greater the collection, the greater the number of *type v* errors that can be expected.

To illustrate this effect, consider a search devised according to the four-step technique recommended in a leading legal research text.²⁷ The first step is to identify all of the elements of the search problem. If there are n elements, they can be identified thus:

$$E_1, E_2, \dots, E_i, \dots, E_n.$$

Associated with each of these elements are two numbers that correspond to the recall and fallout of single-element, full-text searches using that element. The proportion of relevant documents using element E_i , which would be the recall of a search using only E_i , can be denoted E_ir . The proportion of irrelevant documents that use E_i corresponds to the fallout of a search on E_i , and will be denoted E_if .

Adopting some simplifying assumptions,²⁸ the recall for a search using a combination of elements is equal to the product of the recall values for the individual elements. This may be expressed thus:

$$S(E_i, E_j, E_k)r = E_ir \cdot E_jr \cdot E_kr$$

Fallout is calculated the same way:

$$S(E_i, E_j, E_k)f = E_if \cdot E_jf \cdot E_kf$$

Assume that for all elements, recall is 70 percent and the fallout is 5 percent. Assume also that the researcher is willing to put up with as many as twenty-five irrelevant cases. In a collection of 500 documents, 5 percent fallout will yield twenty-five false drops, so only one search element need be used, and recall will be equal to the single-element recall, 70 percent. As the data base grows, however, additional search elements will be needed. To limit the number of false drops in a 10,000 document data base, two elements are needed, and recall drops to 49 percent. When the data base contains 200,000 documents, three elements are necessary, and recall drops to about 34 percent.

So far it has been assumed that the searcher has found all of the proper search terms to use, but this is unlikely to be the case. The probability that the searcher will choose any particular element for the search also needs to be taken into account. If it is assumed that the searcher chooses the term

27. J. JACOBSTEIN & R. MERSKY, *supra* note 5, at 438-41.

28. Here it is assumed that the various values of E_ir and E_if are independent of each other. The independence assumption is false for many (if not most) legal research problems, but the purpose of this discussion is not to provide a practical method for calculating recall and fallout, but rather to show that they vary together. A second assumption is that the connectors used to join the elements are simple "ANDs" rather than more sophisticated proximity connectors. It is not good search practice to use AND connectors, see J. JACOBSTEIN & R. MERSKY, *id.*, but allowing for the effects of proximity connectors would add an unenlightening layer of complexity to this example.

for the search with the same 70 percent reliability that was assumed for the use of the term in the text of the desired document, another 70 percent factor needs to be considered calculating the effect of each search element. The recall of a one-element search declines to 49 percent, a two-element search to 24 percent, a three-element search to about 12 percent, and so on.

Since the individual recall factors for elements are usually unknown, it is difficult to give a more concrete and realistic example in which recall can be calculated. Individual fallout factors can be estimated, however, and an example based on these may be helpful.

In real data bases, the fallout associated with individual elements is likely to vary considerably. It will be closely related to the frequency with which the characteristic words of the elements are used in the natural language of the collection.²⁹ The key to maximizing the effectiveness is to pick search elements that are individually associated with high recall and low fallout.

This research question is used to illustrate the query formulation technique in the *Fundamentals of Legal Research*.³⁰ The searcher wants to find cases discussing the legality of a dog's "sniff search" of a high school student's locker. In *Fundamentals*, nine elements are identified, of which three are selected as search elements: "sniffing," "dog," and "school." Using LEXIS's CAL library and CASES file, the relative frequencies of the search elements "sniffing," "dog," and "school" appear as follows.

59,104	= total number of documents
6,027	= number of documents containing "school"
685	= number of documents containing "dog"
75	= number of documents containing "sniff!" ³¹

Ignoring what might be anticipated to be a small number of relevant documents for this search, and rounding (to avoid creating a specious air of accuracy), these approximate values for individual element fallout are obtained:

29. It has been hypothesized that the distribution of words in natural language is roughly in proportion to the terms of the harmonic series, that is, that the Nth most common word in the language occurs $1/N$ as often as the most common word. See G. ZIPF, HUMAN BEHAVIOR AND THE PRINCIPLE OF LEAST EFFORT: AN INTRODUCTION TO HUMAN ECOLOGY (1949). If this is correct, the number of occurrences of common words in a data base increases much more quickly than the size of the lexicon. The most common words have virtually no value as search terms because they are so common, and full-text systems protect themselves against fruitless searching by making many common words part of an unsearchable "stop list." Even after the elimination of the stop list words, however, large full-text systems contain many words that are too common for productive searching.

30. J. JACOBSTEIN & R. MERSKY, *supra* note 5, at 438-41.

31. All of the figures for this example (including related searches in the New Mexico and federal data bases) were obtained from the LEXIS system from searches run in late April of 1984.

$$E_{sniff,f} = .001$$

$$E_{dog,f} = .01$$

$$E_{school,f} = .1$$

The best search element, at least for the purpose of bringing fallout under control, is "sniff!". In some data bases, "sniff!" is a sufficient search. The New Mexico collection of cases in LEXIS, NM—CASES, contains about 6,000 cases, and a search on "sniff!" alone retrieves about six irrelevant cases.

But the CAL—CASES data base is so large that "sniff!" alone will not sufficiently reduce fallout. Recalling that fallout is approximately u/N :

$$u = E_{sniff,f} \cdot N = (.001)(60,000) = 60$$

As a result, there are sixty *type u* errors to be expected from a single element search on "sniff!" in CAL—CASES; sixty irrelevant cases retrieved by the search. This is, at best, a minimally acceptable result.

Assume that the maximum number of irrelevant documents that the searcher is willing to tolerate is twenty-five. The search must incorporate another element to further reduce fallout. This can be done by adding the next element to the search, "dog":

$$u = E_{sniff,f} \cdot E_{dog,f} \cdot N = (.001)(.01)(60,000) = .6$$

But this calculation assumes that "sniff!" and "dog" are independent of each other, and that is certainly not the case. Sniffing is an activity characteristic of dogs, and it may be assumed that the word "dog" is much more likely to occur in an irrelevant document that contains one of the words subsumed under "sniff!" than in a random irrelevant document. In practice, this search retrieves about six irrelevant documents, more than would be predicted if one assumes complete independence, but still a manageable number.

In a still larger data base, more terms are necessary. LEXIS's GENFED—CASES data base is about ten times as large as CAL—CASES.³² To reduce the *type u* errors to a manageable level, it is necessary to add yet another element to the search request, in this case, "school."

The question that can be done with a one-element search in New Mexico requires a two-element search in California and a three-element search in federal courts. This is bound to have an effect on the recall of searches in the larger files. As the single term fallout factors are combined to reduce fallout, a corresponding reduction is taking place in recall.

32. The figure used for the size of the GENFED—CASES data base (600,000) is an estimate based upon limited knowledge of the size of the similar data base in WESTLAW. See *infra* note 38. The author was unable to make LEXIS count the total number of cases in this file.

Measuring the exact effect of file growth on recall in this example is impossible, but it is possible to get some indication of the effect by displaying the results of the various searches graphically. Figure 1 is a graphic depiction of the results of a test of LEXIS searches using the terms "dog," "sniff!" and "school," singly and in all possible combinations, in the data bases NM—CASES, CAL—CASES, and GENFED—CASES. For any search returning twenty-five or fewer cases, each case was examined for relevance.³³ Each case is represented in the appropriate area of the diagram either by an *x* for a relevant case or a *u* for an irrelevant case. The *x*'s correspond to hits, and the *u*'s to *type u* errors. In regions covered only by searches that returned more than twenty-five cases, it was assumed that output overload had occurred, and the cases were not examined. The total number of retrieved documents for these regions is entered in numerals.

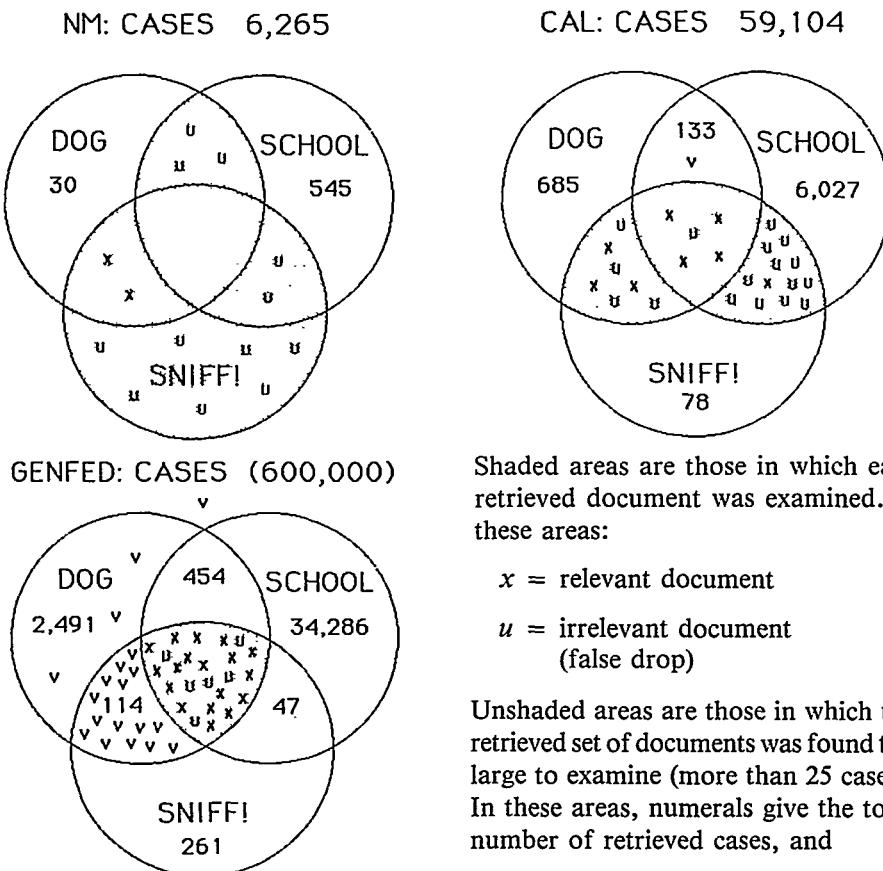


Fig. 1

33. Relevance here was determined by the author. All cases that seemed fairly analogous were considered relevant, not just those sniffing dog cases that were on "all fours" with the hypothetical facts.

Also mapped onto the diagrams in figure 1 are a small number of relevant documents found manually.³⁴ These cases are mapped using the symbol ν , which indicates a *type v* error.

The diagram indicates that the addition of search elements to certain searches resulted in relevant cases being missed. These are most apparent in the GENFED—CASES data bases, in which it was not possible to evaluate the 114 cases retrieved by the search “dog and sniff!”.

VII. The Experiment of Blair and Maron

The first large-scale experimental test of the effectiveness of a full-text information system was performed by David Blair and M.E. Maron.³⁵ The test data base was a group of almost 40,000 documents in a full-text retrieval system that was created to provide litigation support in a complicated lawsuit. The system was evaluated by having lawyers who were working on the case submit sample search questions. The actual searching was done by paralegals,³⁶ and the results were evaluated by the requesting attorneys.

In the experiment, no question was considered answered until the attorneys were confident that the search had retrieved at least 75 percent of the total number of relevant documents. Further investigation showed, however, that while the average precision was 79 percent, the average recall achieved by the system was no better than 20 percent.

One of the most surprising features of this result is that it is so surprising. Surely, one would think, before investing millions of dollars in the existing legal research systems, *someone* should have made at least a rough determination as to what the recall performance of such systems was going to be. But conducting document retrieval experiments is subject to many theoretical and practical difficulties and is expensive. The experiment undertaken by Blair and Maron took over six months to complete and had a total cost of almost half a million dollars.³⁷

The primary problem in determining recall is in finding the total number of documents relevant to sample search requests. In many of the early experiments, a small special data base was constructed. Don R. Swanson's seminal study of a full-text retrieval system³⁸ was based on sample searches

34. Manual research was limited to an examination of all of the cases cited in Annot. 31 A.L.R. Fed. 931 (1977) and its October 1984 pocket part. The topic of this annotation, “Use of Trained Dog to Detect Narcotics or Drugs as Unreasonable Search in Violation of Fourth Amendment,” might be expected to cover all cases directly on point, but not all other relevant cases. This bias accounts for the concentration of unfound relevant cases in the GENFED—CASES “dog” and “sniff!” search.

35. Blair & Maron, *supra* note 2.

36. The study included a smaller test in which the requesting attorneys did the actual searching. *Id.* at 294-95.

37. *Id.* at 298.

38. Swanson, *Searching Natural Language Text by Computer*, 132 SCIENCE 1099 (1960).

in a collection of only 100 documents. Other key papers on full-text retrieval³⁹ drew their conclusions from tests on sample data bases of 273 and 450 documents.

But these small data bases do not have the same performance characteristics as the large full-text data bases used in CALR systems, and so the results of these early experiments are not trustworthy guides to the performance of the current systems. For example, the ALLSTATES data base on WESTLAW contains over one million documents.⁴⁰ Even if the user of such a system were willing to look through as many as 100 irrelevant documents, fallout for the system would have to be limited to about 0.005 percent. This rate of fallout applied to a data base of 450 documents would yield a little more than one-fiftieth of an irrelevant document. Fallout levels this low simply cannot be accurately measured in tests on a data base this small.⁴¹

Another difficulty in measuring the effectiveness of document retrieval systems is in finding the total number of relevant documents for the sample search queries. It is impractical in any large data base for the requestor to evaluate the relevance of every document in the collection, so there needs to be some other means of finding relevant documents. One key study that compared various conventional indexing techniques⁴² analyzed 1,200 documents with respect to 280 search questions. This study relied chiefly on a comprehensive screening of cases for relevance by five postgraduate students, but subsequent study has suggested that the project identified only about 10 percent of the true total number of relevant cases.⁴³

Since there has been so much uncertainty in previous tests of retrieval effectiveness, it is prudent to subject the test performed by Blair and Maron to careful scrutiny before accepting its results. The methodology of the study is well described in the article itself, but certain key issues will be discussed here.

39. Salton, *A New Comparison between Conventional Indexing (MEDLARS) and Automatic Text Processing (SMART)*, 23 J. AM. SOC'Y FOR INFORMATION SCI. 75 (1972); Salton, *Automatic Text Analysis*, 168 SCIENCE 335 (1970).

40. To get some indication of the size of the WESTLAW data bases, the author ran the search "BANC COURT MEMORANDUM TRIAL CASE LAW JJ JUDGE JUSTICE PER" in ALLSTATES on April 19, 1985. The search returned 1,066,550 cases. A similar query in ALLFEDS was aborted by the system after some 431,000 cases had been found.

41. For a critical appraisal of several of the seminal document retrieval experiments, see Swanson, *Information Retrieval as a Trial-and-Error Process*, 47 LIBR. Q. 128 (1977).

42. C. CLEVERDON, J. MILLS & M. KEEN, *FACTORS DETERMINING THE PERFORMANCE OF INDEXING SYSTEMS* (1966) (referred to in the literature as Cranfield II).

43. Swanson, *Some Unexplained Aspects of the Cranfield Tests of Indexing Performance Factors*, 41 LIBR. Q. 223 (1971).

A. Comparing STAIRS to LEXIS and WESTLAW

The full-text document retrieval system evaluated by Blair and Maron was IBM's STAIRS/TLS.⁴⁴ STAIRS is a system that can be used to handle many kinds of document retrieval problems and has been implemented for legal research⁴⁵ as well as litigation support. STAIRS permits the use of lexical same paragraph and same sentence connectors, and can rank retrieved documents in a number of ways, including by date and by the frequency of selected search terms.

STAIRS/TLS has one useful feature that LEXIS and WESTLAW do not share. The TLS portion of the program can be used to specify a variety of relationships between search terms such as "synonymous with" and "narrower than." Enhancements of this sort have been found to improve the performance of full-text systems substantially.⁴⁶ STAIRS/TLS is a state-of-the-art full-text system, apparently equal or superior to the systems in use for LEXIS and WESTLAW.

B. Method Used to Find Relevant Cases⁴⁷

All of the cases that were considered relevant for the purpose of calculating recall were somehow coaxed out of the STAIRS system. After the requestor had announced that he or she was satisfied that at least 75 percent recall had been achieved, Blair and Maron set to work to find additional relevant documents using two techniques. The first method used was to improve the existing searches by adding synonyms to them. As recounted in the article, the authors exercised considerable ingenuity and perseverance in tracking down equivalent terms.⁴⁸

The second method used was to run searches with fewer elements than the searches used by the lawyers and paralegals. For example, if the lawyers used a search in the form:

A and B and C and D

(where A, B, C, and D each represent elements of the search), Blair and Maron would run searches in the following forms.

44. An acronym for Storage and Information Retrieval System/Thesaurus Linguistic System.

45. The state of New Mexico implemented a STAIRS-based legal research system containing *New Mexico Statutes Annotated* and recent New Mexico appellate decisions. The author made use of this system in 1979 and 1980 in his capacity as a judicial clerk and found the operation of the system functionally equivalent to that of WESTLAW.

46. See Swanson, *supra* note 38.

47. For this and the following point, the discussion is based in part on the author's correspondence with David Blair, who supplied information not in the published account of his experiment.

48. Blair & Maron, *supra* note 2, at 295-96.

- A and B and C
- A and B and D
- A and C and D
- B and C and D

Both of these techniques are recommended by CALR authorities as part of ordinary query formulation,⁴⁹ but the increase in the total number of relevant cases found by Blair and Maron is not simply a reflection of their greater skills in query formulation. The expanded search for relevant cases consumed a tremendous amount of time and effort. In some cases, the computer time alone for searching a question amounted to more than forty hours.

Even if the shortcomings of the search result could be attributed to poor query formulation, it is unreasonable to expect the average user of the system to have the expertise of Blair and Maron. The lawyers and paralegals who were the primary searchers were of at least average skill in the use of full-text retrieval systems. Such systems are designed and marketed to be easy to use.

It is important to note that Blair and Maron do not claim to have found all of the relevant documents in the data base. There were almost certainly some relevant documents that were not found by anyone. Since these documents were not taken into account in calculating recall, the true recall of the system is bound to be somewhat less than the 20 percent cited in the study.

C. Performance of STAIRS with Respect to Relevant Documents

When the documents were judged for relevance by the requesting attorneys, they were divided into four categories: irrelevant, marginal (M), satisfactory (S), and vital (V). It might be thought that the system was seen as performing poorly primarily because it failed to turn up many marginally relevant documents. In the basic calculation of results, all three categories (M, S, and V) were considered relevant for the purpose of calculating recall. If the threshold level for relevance had been raised to S or V, results would be as shown below.

Document Categories

Considered Relevant	Recall	Confidence (95%)		Precision
1. V + S + M	19.99%	+ or -	4.95	78.97%
2. V + S	25.30%	+ or -	6.60	56.58%
3. V only	48.24%	+ or -	14.80	18.22%

The performance of the system in retrieving the most relevant cases seems

49. See, e.g., J. JACOBSTEIN & R. MERSKY, *supra* note 5, at 438-41.

to be somewhat better than its overall performance. The improvement between the first and second categories is not significant at the .05 level, but the 48 percent success of the system at finding the "vital" documents does represent a significant improvement. Relative to the needs of the lawyers using the system, however, the improvement is small. The lawyers using the system, who were generally willing to accept a recall level of 75 percent for the system, stated that they must have 100 percent of the "vital" documents to be able to try the case. The 28 percent improvement in the retrieval of these vital documents is not enough when considered in the context of the increased need to find them.

D. Similarities between Litigation Support Data Bases and CALR Data Bases

It is not clear whether the differences between litigation support applications and CALR applications for full-text systems are significant. A litigation support data base is likely to have a greater variety of kinds of documents in it. The system studied by Blair and Maron contained many kinds of documents, including reports, verbatim transcriptions of meetings, correspondence, and memoranda. The style of these documents varied between formal and slangy, technical and nontechnical. The sorts of documents contained in CALR systems are more uniform in tone and format, and thus it might be expected that the problems of synonymity in a case law data base would be less severe than in a litigation support collection. For example, a child might be referred to in a case as a "child" or a "minor," but probably would not be called a "punk" or a "rug rat." With the documents to be searched more or less limited to a formal vocabulary, it might be expected that fewer cases would be missed by the searcher.

But the relative uniformity of the vocabulary in case law is not an unmitigated blessing. Full-text systems perform best when they are looking for distinctive words. But the more often a relatively small number of formal words are used, the less distinctive each of those words becomes. The effect is similar to that of file growth because each term must have, on average, more locations in the collection.

Litigation support systems like the one studied by Blair and Maron are devoted primarily to searching for the facts that might be important at trial. As discussed above, CALR systems are at their best when they are looking for fact patterns because the words that are used to describe fact patterns are more distinctive than those used to discuss legal theories or procedural matters. On this point, then, it might be expected that full-text litigation support systems would perform better than CALR systems.

The question of the applicability of the findings of Blair and Maron to CALR systems cannot be settled without another experiment. But for the time being, the similarities between the two kinds of systems appear to be much greater than their differences. An analogy might be made to a problem frequently encountered in the testing of drugs: a drug has been shown to cause cancer in pigs, but no tests have been done with human subjects. The only prudent thing to do is to assume that the drug is also dangerous to people until proven otherwise. The findings of Blair and Maron may not prove conclusively what level of recall may be expected of full-text CALR systems, but there is no longer a complete lack of persuasive analogy. The proponents of full-text searching now bear the burden of showing that the finding does not apply to CALR.

VIII. WESTLAW and "Full-Text Plus"

The full-text searching experiments done by Blair and Maron show that there are some basic weaknesses in full-text document retrieval systems, and in their analysis of their search results they make a convincing argument that these shortcomings are due to the lack of indexing and authority control. In WESTLAW's case law data bases, it is possible to search manually prepared abstracts as well as the original text of the case. This portion of the paper will consider the extent to which the editorial apparatus attached to WESTLAW cases improves the retrieval performance of that system.

One recent study shows that the editorial apparatus attached to cases in WESTLAW results in about 25 percent more cases being retrieved.⁵⁰ It is not clear, however, what proportion of these additional cases were relevant cases (tending to improve recall), and what proportion were irrelevant (tending to worsen fallout). The study concludes that at least some of the additional retrieved cases were relevant. An examination of the nature of the editorial additions available on WESTLAW, however, suggests that the improvement in recall is probably minimal.

Two kinds of abstracts are added to WESTLAW's full-text case law data bases, headnotes and synopses. The difficulty with each of these editorial enhancements is that they are not suited to computer retrieval.

The limited value of having headnotes in the data base can best be shown by example. In the text of *State v. Baumann*, the following passage occurs:

50. Coco, *Full-Text vs. Full-Text Plus Editorial Additions: Comparative Retrieval Effectiveness of the LEXIS and WESTLAW Systems*, *LEGAL REFERENCE SERVICES Q.*, Summer 1984, at 27.

Here, appellant argues that the securities sold were exempt from the registration requirement by virtue of A.R.S. §§ 44-1843(8) and 44-1843(10). We note that A.R.S. § 44-2033 places the burden of proving the existence of an exemption upon the party raising the defense. [citing authority]⁵¹

In the West digests, this point of law is represented this way:

SECURITIES REGULATION

II. State Regulation (Blue Sky Laws)

(C) Offenses and Prosecutions

. 325. Criminal Prosecutions

327. —— Evidence in general.

Burden of proving existence of exemption from securities registration requirements is upon party raising the defense. A.R.S. §§ 44-1843, subds. 8, 10, 44-2033.

But the representation of this headnote on WESTLAW is this:

349k327

SECURITIES REGULATION

Evidence in general.

Ariz. 1980

Burden of proving the existence of exemption from securities registration requirements is upon party raising the defense. A.R.S. ss 44-1843, subds. 8, 10, 44-2033.

There are several things to note about the various representations of this point of law. The first is that the language used in the text of the headnote is almost identical to the language of the case itself. West's headnoting policy seems to be to track the language of the court as closely as possible. This may help insure the accuracy of the headnote, but it does nothing to enhance the chances of finding the case in a full-text search. Any keyword search capable of finding the headnote given in the example is almost certain to find the analogous passage in the text of the case.

The augmentation of the data base provided by the headnote is not in the text of the headnote, but rather in the subject headings that introduce it. But most of the subject headings are not in the WESTLAW representation of the headnote. The full list of subject headings contains several terms that are not in the passage from the case itself, and so might improve the chance that a searcher would find this case. Since, however, the intermediate headings are not in the searchable document, the enhancement is limited to "Securities Regulation—Evidence in general," which is unlikely to be helpful.

51. 125 Ariz. at 412, 610 P.2d at 46 (1980).

The presence of the first and last subject headings in WESTLAW headnotes also can retrieve irrelevant documents. For example, the final subject heading for Securities Regulation key number 277 is "Renewal, modification, revocation, or suspension."⁵² If the searcher is interested only in cases dealing with renewal, the presence of cases dealing with modification, revocation, or suspension in the same key number is likely to create false drops.

Another way in which the headnotes in WESTLAW can be used to find additional cases is by searching on the key numbers themselves. If the searcher knows that cases of interest are likely to be collected under Securities Regulation key number 327, those cases can be found with a key number search. But this sort of search already can be done quickly and easily in the printed West digests. Searches that combine key number terms with keywords are sometimes valuable, but their effectiveness is limited by the shortcomings of both systems.

The inadequate representation of the key number system available on-line makes it necessary to consult a printed digest in formulating a key number search. As mentioned in section III, the two devices that facilitate the use of West digests are the browsability of the key number system and the presence of the descriptive word index. Neither the outline of the key number system at the head of the topic nor the descriptive word index is available on-line. The dropping of intermediate subject headings in the WESTLAW representation of a case makes it easy for a user of the system to be misled. For example, a searcher who finds the WESTLAW version of the headnote shown above sees only "Securities Regulation—Evidence in general" in the subject headings. Inexperienced searchers may conclude that this key number collects all cases having to do with securities regulation and evidence, not just those involving criminal prosecutions under blue sky laws.

The on-line presence of the synopsis is subject to different considerations. The synopsis is much less likely than the headnotes to track the exact language of the text, and thus is potentially a more fruitful source of enriched vocabulary. But the synopsis, while written in a fairly standard format, is not written in a controlled vocabulary. Further, the synopsis is a short piece of text, and it covers the entire case. Because only the most salient features of the case are mentioned in the synopsis, much of the potentially interesting material in a case cannot be found in the synopsis.

Further, the synopsis summarizes the entire case in a single short paragraph, typically with one sentence for the result in the court below

52. The full analysis of this number is SECURITIES REGULATION; II. State Regulation (Blue Sky Laws); (A) In General; 277 Renewal, modification, revocation, or suspension. Though the headings do not so indicate, it seems to deal with the licensing of securities dealers.

and one sentence for the holding of the case being reported. This condensation brings mentions of very different parts of the case close together, and causes false drops. The computer sees words as being related where they are not.

The addition of good human indexing to CALR data bases is a promising approach to the problem of improving retrieval performance in such systems, but the indexing that is included in WESTLAW is of little assistance. The augmentation of the data base is of a sort that is unlikely to find many additional relevant cases and is likely to produce false drops. Experimental results might prove otherwise, but for the time being it seems likely that "full text plus" improves recall by only a modest amount, perhaps 10 percent. This improvement, applied to the 20 percent recall found by Blair and Maron would bring total recall to 22 percent. Even if the full 25 percent increase in the number of cases retrieved were applied to the text-only recall, total recall would be only 25 percent.

Furthermore, the increase in the number of irrelevant cases found in "full-text plus" searching would exacerbate fallout. This 25 percent increase would have roughly the same effect as a corresponding growth in the size of the collection, and thus cause the same problems discussed in section VI of this paper.

IX. Ramifications for the Users of CALR Systems

Blair and Maron end their paper by alluding to Dr. Johnson's remark: "Sir, a woman's preaching is like a dog's walking on his hinder legs. It is not done well; but you are surprised to find it done at all."⁵³ Should we, in response to the finding of disappointingly poor performance in full-text retrieval systems, abandon the existing full-text CALR systems?

No.

There is an important distinction between the use of full-text retrieval for litigation support and for CALR. In a litigation support application, the document retrieval system needs to be built from the ground up for each case, and it will be virtually the only means of access to the documents it contains. A law firm would be poorly advised to rely on a document retrieval system that can find only 20 percent of the total number of documents applicable to a matter.

But in the context of CALR more than one retrieval system is maintained, and the shortcomings of CALR may be made up by recourse to other research techniques. The question for CALR is not whether it is a sufficient system for legal document retrieval, but whether it makes a valuable contribution to the entire available research apparatus.

53. J. BOSWELL, THE LIFE OF SAMUEL JOHNSON, LL.D. 252 (London 1791).

The Blair and Maron study shows that full-text retrieval systems have a severe limitation. They do not provide comprehensive (or even adequate) retrieval of documents by subject. But they do have many virtues, and they make a positive contribution to an arsenal of research techniques that also includes conventional methods.

The limitations on the effectiveness of full-text CALR are in subject access. Nonsubject legal information needs are often met better by CALR systems than by any other means. Finding cases by docket number, reconstructing partial citations, compiling lists of cases by judge or by attorney, and any number of other applications are ideally suited to full-text retrieval.

Given the limitations that apply to subject searching generally, there are certain kinds of problems for which subject searching by computer is fairly effective. Where the nomenclature of an area of law is settled and the words unusual, it is reasonable to expect full-text CALR systems to provide recall well above the 20 percent average performance. For example, in one question searched by Blair and Maron, recall was 78.7 percent, with precision of 99.0 percent. This was an exceptional case, but there are bound to be exceptional cases, and it is entirely possible that users can be taught to identify such promising cases in advance.

Other situations in which full-text subject searching is useful include cases in which one is looking for a recently invented idea, something too recent to have been incorporated into the controlled vocabulary of conventional indexes. Full-text searching is also useful in finding documents too recent to have appeared in printed indexes.

One must not overlook the usefulness of full-text systems for document delivery. At present, most CALR users print entire documents only when they are otherwise unavailable, usually when they are so recent that they have not appeared in book form. But this limitation is based primarily on the expense of system use, and there is no reason to suppose that the cost of acquiring documents from the computer will always exceed the cost of getting them in books. Another barrier to the use of CALR systems for document retrieval is the absence of standardized page numbering from the existing systems. This also is a problem that is likely to be solved.⁵⁴

The most powerful use of existing CALR systems, however, is in the

54. In June of 1985, Mead announced that it would add "star paging" to its federal case law data bases. Mead planned to embed bracketed page numbers in the text of its cases so that users could tell from the LEXIS display the exact page on which the corresponding material can be found in paper copy. West has sued to prevent Mead from implementing star paging with respect to West's copyrighted publications and, at the time of this writing, has been awarded a preliminary injunction. *West Publishing Company v. Mead Data Central*, 616 F. Supp. 1571 (D. Minn. 1985).

access they give to citations. The advantage of this technique can best be seen by following an experiment. Consider the following search question: "Find all of the cases in which a state court has judicially abolished the defense of contributory negligence in favor of some form of comparative negligence." This question may not be a particularly typical research question, but it is not unreasonable, and it has the great advantage of having a small, known, and well-defined group of cases that satisfy the request.

A moment's reflection shows that this question is not a particularly good candidate for CALR using a subject keyword approach. The words that need to be used as search terms ("contributory negligence," "comparative negligence," "adopt," and so forth) are too common to give the computer a basis for excluding the million cases that did not adopt comparative negligence. Any search that returns a reasonable number of cases is almost certain to exclude some of the relevant cases.

The question might be approached with the following WESTLAW search: "CONTRIBUTORY NEGLIGENCE"/P "COMPARATIVE NEGLIGENCE"/P "ADOPT!"

Run in the ALLSTATES data base, this turns out to be an outstanding search from the point of view of recall because it retrieves all ten of the target cases. Unfortunately, it also retrieves 371 other cases, with the result that, as a practical matter, the search is unworkable. This is a typical example of the output overload associated with file growth. If the ALLSTATES data base contained about 100,000 cases rather than about a million, perhaps the search would have had about thirty-seven false drops rather than 371, and the search would have been manageable. As it is, the search needs to be modified to pare away most of the 371 unwanted cases.

But no plausible narrowing of the search given above seems to retain all ten of the target cases. Adding an element to the search with terms like ABROGAT!, ABOLISH!, and OVERRUL! is disastrous for recall. Courts seem much less likely to write about the negative things they are doing than they do about their affirmative actions. Even changing the same-paragraph connectors to same-sentence connectors loses one relevant case, and it reduces the number of false drops by less than half.

Using full-text retrieval for this problem, then, must give less than perfect recall. Perhaps the best search is to look for the original terms in the digest field. This search finds eight of the target cases with only forty false drops. The searcher who does not know which cases to look for initially, though, is unlikely to end up with the best search.

But to an experienced researcher, this concern with finding all ten of the cases in a keyword search is foolish. All that is necessary is to find one of the target cases, and the others can be located in short order. From the original search, using "age" (inverse chronological) ranking of the out-

put, the twelfth case retrieved is the first of the target cases found. That case cites all of the nine other cases. In general, it is reasonable to assume that each of the target cases cites the cases that came before it and is cited by the cases that came after, so all that is necessary is to find one's way to the end of the string of target cases.

Had the searcher selected "terms ranking"⁵⁵ (output ranked according to the frequency of the appearance of search terms), the second case presented by the system would have been one of the target cases, and the last case (with the full list of citations to the others) could easily have been found by shepardizing.

The citations in reported cases embody a vast amount of research that has already been done by judges, their clerks, and the attorneys in the cases. The presence of citation indexes on-line makes it possible to tap this research and to do much more thorough research more quickly than with any other method.⁵⁶

Shepardization can be done without the aid of a computer, but the implementation of *Shepard's* on-line is particularly convenient. Not only can the researcher find the combined *Shepard's* listing that would appear in all of the applicable volumes and supplements of the paper copy of *Shepard's* in one place, but he or she can also move directly from the *Shepard's* display to the citing cases. Manual shepardization requires so much time, patience, and effort that it is rare for anyone to do it thoroughly. CALR shepardization makes citation chaining fast and easy enough that researchers are more likely to persevere in it.

Finally, it should be noted that while full-text CALR systems make many errors in retrieval, these errors are different from those made by conventional, indexing-based systems. For example, in a small sample study with two search questions, the set of cases retrieved by LEXIS and WESTLAW was compared with the set cited in an *American Law Reports* annotation on the same point.⁵⁷ *ALR* cited fifty-six relevant cases, as com-

55. The benefits of using "terms" ranking appear in this example. All ten of the target cases appear in the first fifty cases retrieved by terms ranking, but "age" ranking buries most of the target cases out of the first 100 cases, with the oldest case ranked as low as 324th.

56. The ease and speed of this technique are attested by the fact that about half of the 400 first-year students trained in the use of CALR systems at the University of Texas Law School in the spring semester of 1985 were able to complete this problem in their first half-hour at a WESTLAW terminal. The students were told to start by finding the first target case in the Missouri cases data base (the Missouri target case is ranked near the top of the output for virtually any plausible search). A reference librarian helped in matters of terminal operation and, to a lesser extent, query formulation.

57. The annotations used are *Offsetting Unemployment Benefits Received against Award for Backpay in Employment Discrimination Actions*, 66 A.L.R. FED. 880 (1984) and *Waiver of Right to Trial by Jury as Affecting Right to Trial by Jury on Subsequent Trial of Same Case in Federal Court*, 66 A.L.R. FED. 859 (1984). The issue covered by the latter annotation has been noted as being particularly ill-suited to CALR. J. JACOBSTEIN & R. MERSKY, *supra* note 5, at 436. This is not a fair comparison between manual research techniques and CALR. The time and effort that goes into the creation of an *ALR* annotation

pared with twenty relevant cases found by LEXIS and twenty-two by WESTLAW. But of the cases found by the CALR systems, twelve were not cited in the ALR annotation, so CALR was the sole source for 18 percent of the total number of relevant cases found in the experiment. This experiment is not large enough to support any quantitative conclusions, but it does suggest that CALR will find cases that are not found by thorough manual research.

This last point is a troubling one, because it threatens to escalate the legal research battle without offering more than a marginal improvement in the substantive quality of the research. A subject search in a CALR system may turn up a few cases that would not otherwise be found, but is it reasonable to expect attorneys to search for such cases when the overall performance of the system for subject searching is so poor?

Consider, by way of analogy, the experience that the law has had with citation verification. It is well recognized that attorneys have a duty to shepardize cases before presenting them in arguments to the court.⁵⁸ But the listing of citations in *Shepard's* citators is ordinarily several months after the date of the citing case. For some time, the only ones worried about closing this interval were the publishers, West and Lawyers Cooperative, each of which established a computer-based system to provide more current information. Recently, however, each publisher has made its current citation verification system available to the public in the form of Insta-cite (West) or Auto-cite (Coop). It remains to be seen whether the use of these systems will be included in the standard of care required of attorneys generally, but at least some practitioners have adopted the policy of using them to check every citation used in an argument. The result may be that legal research has become a more expensive operation for those attorneys. The burden of this extra expense is ultimately carried by the client, and the effect is to make the administration of justice in this country even less affordable than it is now.

This extra expense (if it is one) may be entirely justified in the area of citation verification. The citation verification systems are relatively fast and inexpensive, and their use may save as much time and money as it consumes. But it is not at all clear that the use of CALR systems is justified in all of the cases in which it might turn up otherwise unfindable documents. And yet if other attorneys are going to run comparatively unproductive

is much greater than anyone might be expected to give a CALR search. For a checklist that shows the depth of research that goes into an ALR annotation, see the last leaf of *What is the Difference between Owning Lawbooks and Owning a "System" of Legal Research?*, a promotional booklet distributed by the Lawyers Cooperative Publishing Company.

58. See, e.g., *Golden Eagle Distributing Corp. v. Burroughs Corp.*, 103 F.R.D. 124 (N.D. Cal. 1984), in which sanctions were ordered against a law firm for failing to cite pertinent subsequent authority in an argument.

CALR subject searches for the purpose of finding a few extra cases, is it safe for any of their potential adversaries not to do so?

This is a dilemma. In some applications, CALR is a better, faster, and cheaper tool to use, and the administration of justice in this country can be expected to be better for its existence. But in other applications, full-text CALR may become a burden on everyone. What it does poorly it does just well enough that it cannot be ignored in an adversarial setting.

X. Conclusion

An examination of the findings of Blair and Maron with respect to the recall performance expected of full-text retrieval systems does not indicate that existing CALR systems should be abandoned. It does suggest, however, that they are severely limited in their usefulness for certain applications, and that there is a good deal that could be done to improve them. A detailed proposal for the improvement of CALR systems is beyond the scope of this paper. As an introduction to a future paper, however, a number of suggestions seem worthy of fuller examination.

The primary lesson of the Blair and Maron study is that keyword searching on the full text of a case does not ordinarily provide adequate subject access to that case. The computer simply does not have the intelligence to distinguish relevant cases from irrelevant cases on the basis of their full texts alone. If the systems are to be improved, the most obvious step is to add good human subject indexing to the data bases.

The intelligence of the processing algorithm for full-text CALR systems also could be improved. In particular, the systems automatically could search for words related to the search terms provided by the user. Both LEXIS and WESTLAW now automatically search for plurals and a few other equivalent forms, but the systems could be made to search for less obvious synonyms also, or at least to suggest alternative terms to the user.

The systems also could offer more help to the user in choosing search elements. Since the approximate fallout associated with individual search elements is related to the frequency of use of the words that embody the element, a system that gave the user information about the frequency of words in the data base could be used to improve search performance. Perhaps a system could be built to accept a full list of elements from the user and automatically formulate a search based upon the most distinctive terms.

This latter capability could be combined with a move away from the strict boolean rules for the inclusion and exclusion of retrieved documents toward a more flexible approach. While the searcher was listing search elements for the machine, he or she also could provide weighting factors that would help the system calculate the importance of each individual

term to the query. The final ranking of documents retrieved by the search could be determined by reference to this weighting.⁵⁹

Finally, much more could be done to facilitate the computerized searching of citation chains. Even with the vast improvements that have been realized by the present computer implementations of legal citation indexes, there is a great deal of room for improvement.

Nothing, however, can really lift the curse of Thamus. Written records are, in many ways, incurably obscure representations of ideas:

[T]hat's the strange thing about writing, which makes it truly analogous to painting. The painter's products stand before us as though they were alive, but if you question them, they maintain a most majestic silence. It is the same with written words; they seem to talk to you as though they were intelligent, but if you ask them anything about what they say,...they go on telling you just the same thing forever. And once a thing is put in writing, the composition, whatever it may be, drifts all over the place, getting into the hands not only of those who understand it, but equally of those who have no business with it; it doesn't know how to address the right people, and not address the wrong.⁶⁰

59. It is ironic that some of these suggested improvements are similar to features of the WESTLAW system that have been abandoned. WESTLAW used to have an on-line listing of all searchable words contained in its data bases, together with figures showing the frequency of appearance of each. WESTLAW also used to promote the use of a nonboolean search logic that accepted lists of search terms and ranked the output according to the number and frequency of the occurrence of the search terms. Finally, WESTLAW also used to depend entirely upon human abstracting (in the form of headnotes) for retrieval. An examination of the reasons for the changes in WESTLAW would be instructive.

60. PLATO, PHAEDRUS 275d-e.