The Electronic Searching of Law

Reed Dickerson

Indiana University School of Law

Follow this and additional works at: http://www.repository.law.indiana.edu/facpub

Part of the Computer Law Commons, Legal Writing and Research Commons, and the Library and Information Science Commons

Recommended Citation

http://www.repository.law.indiana.edu/facpub/1503

This Article is brought to you for free and open access by the Faculty Scholarship at Digital Repository @ Maurer Law. It has been accepted for inclusion in Articles by Maurer Faculty by an authorized administrator of Digital Repository @ Maurer Law. For more information, please contact wattn@indiana.edu.
The Electronic Searching of Law

Although the use of electronic computers in legal research has not yet been perfected, Professor Dickerson believes that computers may be at least a partial answer to the problem created by the ever-increasing mass of statutory material and judicial decisions that must be consulted to determine the law on a given point. He explains just how the computers are used in legal research and points out their limitations as well as their advantages. [For a general discussion of this subject, the reader is referred to the article, “Prepare Now for Machine-Assisted Legal Research”, by Roy N. Freed, 47 A.B.A.J. 764 (August, 1961).]

by F. Reed Dickerson • Professor of Law at Indiana University

It is possible that the 83d Annual Meeting of the American Bar Association, held last year in Washington, D. C., will be remembered best for a series of demonstrations held under the joint sponsorship of the Association’s Electronic Data Retrieval Committee,1 The Health Law Center of the University of Pittsburgh, The International Business Machines Corporation and the United States Patent Office. These widely differing organizations joined forces to put on one of the most unusual legal exercises of recent years. This was the first public demonstration of the use of modern electronic computers to search both statutes and case law.

Why was this so important for lawyers generally?

Two hundred years ago, it has been said, a lawyer could carry in a wheelbarrow almost all the law books he needed to consult in his daily practice. Today the same wheelbarrow would hardly hold all he needs to consult with respect to just one law, the Internal Revenue Code. A fund of case law estimated at 2,200,000 reported opinions (increasing at 25,000 a year), 77,000 key numbers to reckon with, and 2,000,000 entries in descriptive-word indices, all testify that the flood is already here. The rising tide of statutes and case law is forcing today’s lawyer to ask, almost anxiously, what can be done to cope with it.

Unfortunately, traditional methods of search are breaking down, not only because they are inordinately time-consuming in many cases but because they are increasingly inadequate to uncover the relevant materials the lawyer is looking for. This in turn undesirably increases the element of chance in the trial of cases and the planning of business affairs.

The solution would seem to be to make more accessible and manageable the multitude of statutes and case materials that cannot be eliminated by persuading the legislatures and courts to remove needless complexities or by publishing a smaller proportion of the total number of judicial opinions.

The general problem is, of course, not peculiar to the law. The burden of storing and finding information appears to be a chronic problem in our culture for which there is no single or simple solution. What follows is a description not of final solutions but a progress report on several specific approaches that have been successfully tested and that, for some kinds of legal problems, promise to be economically feasible. They are described not only because they work, but because they offer advantages not previously offered by most comparable attempts in non-legal fields. Some of these advantages are discussed below. Because of the pitfalls and unsolved problems still remaining, any claim that the vast materials of the law are at last at the lawyer’s finger tips would be highly premature.

The following method of searching statutes was developed for the most part at the University of Pittsburgh as an outgrowth of the recent project of its Health Law Center to develop a comprehensive compilation and analysis of state laws affecting the operation of hospitals. Dissatisfaction with traditional methods of searching statutes led the Director of the Center,2 to explore the possibility of adapting existing methods of electronic storage and retrieval to the searching of statutes. The result was a system that will now be described in terms of a legal problem in the field of health law.

Suppose you are a lawyer for a hospital association. Suppose your association wants to develop a kind of metropolitan hospital the first floor of which can be used for drug stores, beauty shops, and other small retail enterprises. A legal problem arises because the association does not want the hospital to lose its entire tax exemption (it hopes that its tax will apply only to the first floor). Unfortunately, the local state statute does not permit split listing of this kind, and the association, which is considering proposing a bill to the legislature, wants to know

1. Under the chairmanship of Reed C. Lawlor. Of Los Angeles.
2. John F. Horthy.
what other states' statutes have to say on the subject.

To solve such a problem by traditional methods, you would have to go to a library having the statutes of all fifty states and manually search them. This would probably take many hours and, in view of the limitations of indices and legislators' tendency to tuck particular provisions into unlikely nooks and crannies, you would not be sure even then that you had found all the relevant provisions. You would spend a lot of time looking in some states for statutes that weren't there.

Contrast with this the method of statute searching demonstrated at the American Bar Association meeting. But to understand how the material is found electronically, it is necessary first to understand how it is stored.

Output of a Machine Depends Upon the Programming

The first principle of machine retrieval is that, while it may be an overstatement to say that nothing can be obtained from a machine that has not been put into it, it remains true that the output of a machine depends on its programming and its input. In this case the machine was a general purpose IBM 650 magnetic tape system and stored on one of its tape reels was the text of all the state statutes affecting hospitals. (Ideally, the machine should have contained the state statutes on all subjects, but lack of funds limited the demonstrations to a particular class.)

The statutes were put into the machine in this way: First the citation and entire text of each statute were typed on punched cards. It took many cards to carry each statute. Each time the key punch operator typed a letter or figure at the top of the card, the machine punched two rectangular holes below the letter or figure in an arrangement that represented that character. When the stacked cards were fed into the machine, the holes triggered the electronic impulses that translated the individual characters of the text into magnetized spots on a revolving reel of magnetic tape. The completion of this operation produced the "text tape".

By an electronic sorting process that few lawyers will understand (or need to understand beyond its specific capabilities and limitations), the machine next produced a "vocabulary tape". This is a comprehensive word concordance, an alphabetical index of the words actually used in all the statutes on the text tape. For instance, all uses of the word "partnership" were collected and each use was identified by a number representing the particular section of the statute in which it appeared. The entries for "partnership" and its immediate neighbors appear like this on the print-out:

\[
\text{PARTIES} \quad 66 \quad 122 \quad 399 \\
\text{PARTLY} \quad 368 \quad 391 \quad 412 \quad 426 \\
\text{PARTNERSHIP} \quad 170 \quad 270 \\
\text{PARTNERSHIPS} \quad 66 \quad 77 \\
\text{PARTS} \quad 128 \quad 362
\]

Thus, the word "partnership" appears in two different sections identified by the numbers "170" and "270". While such a vocabulary list appears to be an exhaustive and therefore a wholly unselective index of every word used in the statutes, it omits a number of common words that have little or no search value. These are words such as "a", "an", "between", "by", "even", "every" and "from". Otherwise, every word is catalogued alphabetically.

Unlike most storage-and-retrieval systems, this system uses no coding and no selective system of classification. There is no table of contents, no selective index, and no keyed classification system interposed between the searcher and the raw materials stored on the text tape. This means that the use of the system, unlike several to be described later in this article, need not await the development of an acceptable standard vocabulary or selective classification system. It means also that, wanting the advantages (and limitations) of an indexer or classifier, it is incumbent on the searcher to ask the right questions and couch them in the right way.

Despite their thoroughness and agility, these machines are, as IBM adviser H. Peter Luhn has put it, "incredibly stupid". For one thing, being limited to the manipulation of symbols, they are wholly incapable, without specific instructions, of dealing with language difficulties such as those posed by ambiguities or synonyms, or even such minor word variations as misspellings. Accordingly, the searcher must be careful to envisage every possible word variation in which his problem may be couched in the statutes. This is not an unsuperable burden if the searcher works, as he must, from the printed vocabulary list and if the relevant vocabulary is not so heterogeneous that it is inordinately long. The only barriers between him and his materials are the limitations of his own imagination and ingenuity, limitations that are equally significant in traditional search methods.

The printing of this vocabulary list completes the storage phase of the storage-and-retrieval process. Of the two principal phases it is by far the more expensive and time-consuming. In terms of traditional search methods, this phase is to be compared with the writing, editing, printing and publication of a state code of statutes. Even here, the comparison is favorable to machine methods. Apart from the initial cost of the machines themselves, the big cost lies in the labor involved in transferring the text of the statutes to punched cards, a cost that can be drastically lowered by using punched tape that is automatically produced when the statutes are initially printed and ultimately by using some kind of electronic scanner capable of transferring printed words directly to tape.

We are now in a position to ask for the answer to our split-listing problem.

The Crucial Step Is Preparing the Inquiry

The crucial step in machine retrieval is to prepare an adequate inquiry. For this purpose the machine itself is of no help. Moreover, the split-listing problem cannot be adequately searched in terms of "split listing" because that phrase is a colloquialism seldom used in statutes. Instead, we must use for our main search a more conventional legal terminology. This, of course, must be drawn solely from the vocabulary list.

The legal problem involves three basic concepts. The first is that of "tax". The second is that of "exemption". The third is that of "partly commercial". The technical problem is
to retrieve each state statute that deals with all three concepts.

Here we meet the problem of synonyms, synonymous expressions, and alternative methods of description not involving the use of synonyms. Because not every statute deals with the tax concept in the same way, our inquiry must reflect every possible way in which that idea has been expressed. Working from the printed vocabulary rather than the statutes, we select the following key words (technically known as "descriptors"):

**Taxation**  Tax  Taxed  Taxable

**Similarly for the second concept:**

Exemption  Exempt  Exempting  Exemptions

**And for the third:**

Portion  Partly  Commercial

Now we ask the machine to list all the state statutes that use any word from the first bundle of descriptors, any word from the second, and any word from the third. Mechanically, this is done by putting the question on punched cards and placing them in the machine, in the same manner as the original text.

The machine can answer three kinds of questions. First, it can tell us how many provisions meet the requirements of the question. This may be important, because a very large number may show that the question has been formulated too broadly. Second, the machine can give us the citations of all the statutes covered by the question. Third, it can print the text of those statutes. Whether we ask for complete text depends on the reasonable availability and usability of the text in other forms.

Figure 1 shows the results of asking for the citations and text of the statutes relating to the split-listing problem. After a restatement of the initial inquiry in code form, it shows all the citations and the first part of the text. The time taken to insert the inquiry cards, process the question, and print the answer requires twenty-six minutes in this instance, as compared with the many hours of research that a manual state-by-state search would involve. More advanced machines could do the same job much more quickly.

Figure 1 shows that the inquiry has also produced the case of *State Tax Commission v. Commercial Realty Co.* The reason is that the text tape also includes the relevant case law and this particular case happens to meet the requirements of the inquiry ("tax" appears in the name of the plaintiff, "commercial" in the name of the defendant, and "exempt" in the case abstract). Rather than indicate a flaw in the system, it points up the fact that the machine is not intended to make the final selection but only to produce a manageable fund of promising materials (all that are responsive to the specific question), from which the searcher makes the final selection. Of the nineteen statutes cited by the machine some may turn out, on inspection, to be of little significance to the split-listing problem. This is as it should be, because the machine is designed not to replace legal judgments, but to reduce the kind of mechanical drudgery that lawyers need to avoid if they are to serve their clients adequately. Rather than eliminating the necessity of thinking through the basic legal problem, it highlights the need for careful analysis, not only in framing the inquiry but in evaluating the results.

The great advantage of the machine is not alone in the time it saves but in its ability to do a more thorough and exhaustive research job than is otherwise possible. This ability is not subject to the limitations imposed by inadequate indexing or the booby traps of inept arrangement.

There is little question that from a technological standpoint the system works. If the source materials are on the text tape and if the searcher does a resourceful job of framing his inquiry on the basis of the words appearing in the vocabulary print-out, the machine will produce the results. The

3. Although "partly" and "commercial" are plainly not synonymous, a spot check of the hospital statutes showed that in this particular context these are alternative methods of description.
What About a "Cascade of Irrelevant Materials"?

Fears have been expressed that the use of inapt or too few descriptors in framing the inquiry may produce a "cascade of irrelevant materials". This can readily happen. But must an inquiry be inapt? Once the searcher has mastered the technical idiosyncracies of the machine, which are not extraordinarily difficult, there is no indication that he needs more than the normal legal sophistication necessary for conventional research. He may need less. Even under conventional methods the searcher is potentially confronted with an enormously larger body of irrelevant materials than the machine is likely to supply and his ability to cope with it ultimately depends on how much legal sophistication he brings to the problem in the first instance or how much he can soon acquire by sampling his materials. Fortunately, too, the machine makes it possible to minimize the risk of inundation: It can quickly count and report the number of available items that would be produced by pursuing a proposed question.

Doubts have been expressed that any demonstration that is limited to a series of carefully planned exercises can be convincing, because such an exercise necessarily involves some card stacking or bootstrap pulling. These doubts might have been well founded had the demonstrations been so limited. Fortunately, they also included a number of free searches based on questions framed spontaneously by attending individuals. Incidentally, some of these dramatically illustrated the points made in the preceding paragraphs.

Doubts such as these lead to the more basic question of how the demonstrators could have known how to select the relevant descriptors without first reading all the statutes. Could they have worked merely from the vocabulary print-out? Naturally, it would have been difficult and perhaps impossible to frame significant inquiries without some legal knowledge beyond a mere acquaintance with the vocabulary list. Familiarity with at least some of the hospital laws would be a normal prerequisite. However, this is no more than a researcher using conventional methods would have to bring to the problem, or soon acquire, before he could bypass the "cascade of irrelevant materials" and find what he is ultimately looking for. Moreover, being conversant with typical hospital laws is a far cry from having to read them all. Some conventional research is undoubtedly necessary to reaching significant results under any method, machine or otherwise.

Who can own such a machine? The machine demonstrated costs about $600,000 to buy or about $12,000 a month to rent. Very few law firms or clients can afford this kind of investment. On the other hand, what law firm or client can afford to undertake the costs of writing, editing, printing, and publishing a work such as Me-Kinney's New York Statutes? Fortunately, it is unnecessary to own such a machine to enjoy its benefits. A number of universities and governmental units already have computers, and it is conceivable that local bar associations, legal foundations and private publishing houses will be able to acquire suitable equipment and make it available at hourly rates.

Although much of this is still speculative, the ultimate test of feasibility would seem to be this: Will the costs of maintaining such a machine make possible use-rates comparable to or lower than the cost-rates to the lawyer of doing the same research by traditional methods? Only experience can give the final answer. Although optimum results depend on keeping the machine almost fully occupied, so far this has been no problem, because the machines are used also for solving problems in many fields other than the law.

What kind of legal problem warrants machine treatment? Presumably, it is one having one or more of the following: great importance, particularly financial; voluminous source materials; source materials unpredictably scattered among many jurisdictions; the need for exhaustive treatment; and the need for quick treatment.

The Problems of Searching Case Law

For many lawyers the problems of research are even more serious for case law than they are for statutes. What about the feasibility of searching case law electronically?

Technologically, cases could be handled in the same way: Put the text of the opinions on a series of text tapes, build a comprehensive vocabulary list, and then search the materials in the way already described. The difficulty here is that the aggregate text of the reported cases is vastly more voluminous and heterogeneous in language than that of the corresponding statutory materials. In fact, the time and expense of preparing punched cards made it impractical to attempt to record, for use in the American Bar Association's recent demonstrations, the original text of even a sample of court opinions. Moreover, even if
punched tapes or electronic scanners were practically available for the purpose, it is not as yet certain that it would be desirable to attempt to store on tape the text of the vast body of case law. For one thing, the great diversity of subject matter and the even greater variations in language would, in view of the extent of the materials, mean that the working vocabulary list would approach the dimensions of a comprehensive general dictionary. Whether the average lawyer would be better off with such a system or with one such as will now be described is difficult to determine.

It has been suggested that the same objections apply also to the statutory output of Congress and the several state legislatures. This is not so. When the vast amount of editorial material is subtracted from a state’s legislative code or compilation, the amount of statutory text remaining is small when compared to the amount of judicial text in the same jurisdiction.

Unlike the statutory search, therefore, the systems of searching case law demonstrated in Washington involved a classification or selective indexing system requiring the mediation of a case digester. This not only greatly reduced the bulk of materials recorded on the text tape but gave the searcher the help of a professional classifier or indexer. On the other hand, the searcher was correspondingly inhibited by any inadequacies in the work of the digester. In this instance, the advantages were believed to outweigh the disadvantages.

Digested materials were supplied in the form of case headnotes (Lawyers’ Co-operative Publishing Company), general case abstracts (patent design law materials from the Bureau of National Affairs and oil and gas materials from the Southwestern Legal Foundation) or case abstracts couched in a specially tailored and standardized vocabulary (my own materials on food products liability). Except for the patent cases, these case abstracts were handled by the techniques already outlined for the statutory search.

The key difference lay in the preparation of the headnotes and abstracts before committing them to punched cards. Because I am more familiar with my own materials, I shall discuss the search in terms of the cases on food products liability. Somewhat similar demonstrations were conducted with oil and gas cases by a representative of the Southwestern Legal Foundation.

Figure 2 shows the form that each of the 260 sample products liability cases took before it was transferred to punched cards. Each case was thus abstracted on a separate sheet containing a special seventy-four-word vocabulary. Such a system provides a fund of case abstracts couched in terms relevant for searching purposes and sufficiently standardized to eliminate the synonym problem.

In each case the only inserted material was the name of the case, the citation and the name of the product. The rest was handled simply by checking the applicable words in the vocabulary. The result was a sheaf of cases that, if not too bulky, could be quickly searched even manually. Thus, I interested in locating all the contributory negligence cases could locate them simply by fixing his eye on the tenth item from the end and flipping the pages.

From here on, the system was the same as that already outlined. The key punch operator typed across the top of a series of cards the name of the case,

### Food Products Liability Case Vocabulary Analysis

<table>
<thead>
<tr>
<th>Name of Case</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yochem v. Gloria, Inc.</td>
<td>134 Ohio St. 427, 17 N.E. 2d 731 (1938)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product</th>
<th>Water</th>
<th>2-314</th>
<th>2-315</th>
<th>2-318</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*Food</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Consumer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Retailer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Retailer-labeller</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Wholesaler</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Wholesaler-labeller</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Manufacturer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Restaurateur</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Sale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Foreign-object</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Unwholesomeness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Latent-defect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Sealed-container</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Trichinosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Staphylococcus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Undulant-fever</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Botulism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Typhoid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Mental-anguish</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Allergy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Individual-weakness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Negligence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Breach-of-statute</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Mens-rea-unnecessary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Deceit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Warranty-contract</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Warranty-tort</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Express-warranty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Sales-Act</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*15(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*15(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*15(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Buyers-selection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Commercial-Code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-313</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2

cards were then inserted and the ma-
Breach-of-statute Warranty
someness Latent-defect Negligence
Consumer Restaurateur Sale Unwhole-
the case illustrated in Figure 2:
appears on the punched cards cut for
checked. For example, the following
terms we were able to produce all the
headnote and each key word was as-
signing a five-digit code number. These
in turn were grouped in a 485-word
vocabulary according to key word, simi-
lar to the vocabulary described
earlier except that it was produced
manually and not by machine. Each
d case was then abstracted by reference
to this vocabulary, which made it pos-
sible to describe the case by an average
of only six key words. Next, a punched
card was prepared to show the code
number of each headnote in which the
key word appeared. These cards were
then fed into the machine for storage
and later retrieval.

The RAMAC machine has fifty ro-
tating discs as its “memory” instead
of the plastic tape used by the 650. Of
these only one and a half were used in
this search. As the Director has ex-
plained,

Information is recorded on these discs
in the form of extremely small differ-
ently magnetized spots. A series of six
of these spots, each of which can be
magnetized in either of two polarities,
arranged in a circular path, will define
a distinctive code pattern which repre-
sents a numeral or a letter of the
alphabet. A reading head which can
respond to the magnetized spots moves
mechanically to a selected disc in re-
sponse to an instruction fed into the
machine from a punched card. The
machine is able to intercompare two
or more lists and to generate a re-
sultant list of five-digit numbers in
response to a question to the file.

The specific differences between this
and the other systems are significant
here mainly because they illustrate the
wide range of alternatives now avail-
able. Again, the significant differences
lie not so much in the operations of the
machines themselves as in the
preparation of the legal materials. As
the Director has conceded, case ab-
stracts based on original text are likely
to be better articulated to the substan-
tive materials than those based only
on headnotes already prepared without
this kind of searching in mind. Fur-
thermore, a system that uses no coding,
such as that used for the food products

The Electronic Searching of the Law

its citation, the name of the product,
and each word that the digester had
checked. For example, the following
appears on the punched cards cut for
the case illustrated in Figure 2: “Yo-
chem v. Gloria, Inc. 134 Ohio St. 427
17 N.E. 2nd 731 (1938) Water Food
Consumer Restaurateur Sale Unwhole-

The Limitations of the
The machines
The early test inquiries came
to grief because they were framed with-
out careful regard to the words appear-
ing in the vocabulary print-out. Others
came to grief because spelling variants
were caught, like other word descrip-
tors, in the conjunctive, when they
should have been caught in the dis-
jointive. In such cases, the machine an-
swered, honestly and accurately, “No
listing”. Inquiries that were phrased
too narrowly produced few or no cita-
tions. Those that were phrased too
broadly produced more than a searcher
could cope with, and thus suggested
the desirability of rephrasing the ques-
tion. All these occurrences underlined
this fact: These machines cannot trans-
cend the materials that are stored in
them and they can answer only the
specific inquiries (however inapt) that
are presented to them. Every signifi-
cant legal judgment must still be made
by a live lawyer. These legal judgments
appear to be substantially the same as
those involved in conventional re-
search.

Despite the initial human failures
and toe stubbing, the machine began to
produce impressive results. In five min-
utes we were able to produce all the
Massachusetts restaurant cases in
which the transaction was held to con-
stitute a sale. Similarly, for the recent
foreign object cases based on res ipa
logitur.

The search of patent cases was con-
ducted separately under the guidance of
the Director of the Office of Re-
search and Development of the Patent
Office. It used a RAMAC 305 and a
somewhat different method of digesting
and classifying the source materials.

The sample used consisted of ab-
stracts for about 222 decisions on de-
sign patent law published in the United
States Patent Quarterly during the past
twenty years. These abstracts were de-
veloped in the following way. Legal
digesters working from case headnotes
first extracted what they considered
the key words. These together with
the names of the judge and the tribunal
constituted the search words. Each
headnote and each key word was as-
signed a five-digit code number. These
in turn were grouped in a 485-word
vocabulary according to key word, simi-
lar to the vocabulary described
earlier except that it was produced
manually and not by machine. Each
case was then abstracted by reference
to this vocabulary, which made it pos-
sible to describe the case by an average
of only six key words. Next, a punched
card was prepared to show the code
number of each headnote in which the
key word appeared. These cards were
then fed into the machine for storage
and later retrieval.

The RAMAC machine has fifty ro-
tating discs as its “memory” instead
of the plastic tape used by the 650. Of
these only one and a half were used in
this search. As the Director has ex-
plained,

Information is recorded on these discs
in the form of extremely small differ-
ently magnetized spots. A series of six
of these spots, each of which can be
magnetized in either of two polarities,
arranged in a circular path, will define
a distinctive code pattern which repre-
sents a numeral or a letter of the
alphabet. A reading head which can
respond to the magnetized spots moves
mechanically to a selected disc in re-
sponse to an instruction fed into the
machine from a punched card. The
machine is able to intercompare two
or more lists and to generate a re-
sultant list of five-digit numbers in
response to a question to the file.

The specific differences between this
and the other systems are significant
here mainly because they illustrate the
wide range of alternatives now avail-
able. Again, the significant differences
lie not so much in the operations of the
machines themselves as in the
preparation of the legal materials. As
the Director has conceded, case ab-
stracts based on original text are likely
to be better articulated to the substan-
tive materials than those based only
on headnotes already prepared without
this kind of searching in mind. Fur-
thermore, a system that uses no coding,
such as that used for the food products
liability cases, would seem to be simpler to set up and operate than one that does. A system based on a simple open-end vocabulary would seem to be easier to develop and, in some areas of the law, easier to use than one using a classification system built around a complicated hierarchy of categories.

The system used for searching the health statutes outlined above avoids these difficulties but, so long as it depends on punched cards, storage of the materials requires a lot of time and expense. It may also turn out to place too heavy a burden on the imagination of the searcher. On the other hand, use of a professional digester or indexer requires, as a bare minimum, the development of a satisfactory vocabulary. It is one thing to develop a vocabulary for a limited field such as food products liability or oil and gas. It is quite another to develop one for each field and then blend them into a reasonably consistent whole, which ultimately is the only satisfactory solution. Unfortunately, the existing systems appear either to be unnecessarily heterogeneous or to have been developed under conditions that require their developers to spread themselves too thinly across the face of the law and thus into areas where they were not always sufficiently sophisticated. For the time being, the best approach appears to be to experiment further in limited areas. If an adequate job can be done on these, there is at least no technological reason why it cannot be done ultimately for the law as a whole.

All these experiments point up the necessity of a common language for identification and comparison. However much the use of synonyms and other variants may enrich the color and beauty of language generally, they are only impediments in the scientific development of this phase of the law, once an adequate expression has emerged.

The Potentialities of Electronic Searching

The potentialities of electronic searching are almost staggering. Use of these techniques in the searching of patent claims in the field of steroids during the past three years, entailing an average of five minutes a search, has shown that machine searching can be accomplished at lower cost and with much greater accuracy than traditional methods. A machine search of the existing 2,400 patents in this limited field now produces about twice as many pertinent materials as a manual search. In other words, even expert human examiners, working without machines, overlook on the average almost half of what is looked for.

We are also on the threshold of even more startling developments outside the area of mere storage and retrieval. Electronic devices are now available for analyzing the logical structure of statutes and, if a consistent legislative vocabulary can ever be developed, it will be possible to determine electronically, for example, what specific provisions of a complicated legislative structure such as the Internal Revenue Code need to be changed to effectuate a specific substantive proposal. But this is another story. Closer to our immediate concern is the possibility of abstracting case law by machines capable of identifying significant word affinities by the frequency with which particular words appear in proximity to each other.

Many of those who saw or participated in the demonstrations in Washington were overawed if not overwhelmed by a mass of technical data, economic considerations and speculations on their ultimate impact. In fact, it would be easy to conclude that the whole thing is too complicated, too frightening and premature for adoption. Some may even think that this kind of systematization is alien to the creative spirit of the common law.

Fortunately, we do not have to choose between committing ourselves to the wholesale adoption of electronic methods or rejecting altogether the idea of mechanized search. Many aspects of the techniques discussed in this article are already in use. They work. They are practical. They are economically feasible. And there is a lot to choose from. There are full-blown systems such as those demonstrated in Washington. There are less complicated systems. There are some very elementary systems like the kindergarten one I have been using for many years in the field of food products liability. And so we can be highly selective. We can be cautious.

Many of these systems, although still in the developmental stage, have matured sufficiently that we can feel safe in testing them to see to what extent we can use them in searching particular kinds of legal materials. Ultimately the question is no different from that of deciding whether it is desirable to invest in office dictating equipment.

From these tests, I believe we shall discover that, while most specific legal problems will continue to call for more or less traditional methods of manual research, there will be many that are of sufficient difficulty, legal significance and financial impact to warrant the use of mechanical or electronic devices. Fortunately, the lawyer will not have to own or rent a machine to use it.

Another lesson drawn from the Washington demonstrations is that these machines dispense with neither the lawyer nor the application of any kind of essential legal analysis that a lawyer now uses. Nor, on the other hand, do they appear to increase the burden of legal analysis. Properly used, they may enable lawyers to shed much of the deadening drudgery that most of them would prefer to avoid and that in many cases impairs their capacity to concentrate on the more crucial aspects of legal analysis that they ought to be preoccupied with.

Finally, it seems clear that these devices offer no single, simple solution to today's legal searching problems. We have much to learn about them. We must "make haste slowly". It would be high folly if at any stage lawyers or law publishers were to rush to store electronically a lot of inadequately prepared legal materials and then search them with inadequately prepared inquiries. Nothing could more quickly discredit the fine advances that have already been made.

We owe it to ourselves as lawyers and to the public to investigate and selectively evaluate the rich ore that is being mined here. We should do this boldly, carefully and with the conviction that if we do this job well we shall not only make the practice of law more satisfying professionally but help to assure to the legal profession the full public confidence that it deserves to enjoy.