Computer Retrieval of Case Law

Robert A. Wilson

Recommended Citation
Robert A. Wilson, Computer Retrieval of Case Law, 16 Sw L.J. 409 (2016)
https://scholar.smu.edu/smulr/vol16/iss3/2

This Article is brought to you for free and open access by the Law Journals at SMU Scholar. It has been accepted for inclusion in SMU Law Review by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.
COMPUTER RETRIEVAL OF CASE LAW

by

Robert A. Wilson*

I. INTRODUCTION

Perhaps the most time-consuming and difficult part of any law practice is legal research. For many years lawyers, judges, legislators, and legal scholars have been delayed in their efforts because of the laborious task of locating pertinent precedents. Because legal research is a slow and essentially frustrating part of the legal profession, the practicing lawyer's first inclination is to employ someone to do his briefing. Irrespective of whether the lawyer does the research himself or hires a clerk to do it for him, the cost of getting together the legal information necessary to be successful in his chosen profession is a major item of expense. The purchase price of books and the expense of shelving and maintaining them are only the most visible part of that picture. The hours of time which must be spent in consulting indexes, jotting down references, locating and reading them, and finally discarding most of them as "not in point" constitute a great waste of valuable time and highly skilled brain-power.

Three reasons may be suggested why legal research today is so difficult, so frustrating, and so unsatisfying in its results. First, there is the great accumulation of decisions, statutes, and regulatory materials which has been inherited from the past. Of cases alone, there are 2 1/2 million in the reports,\(^1\) of statutory sections, 1 1/2 million,\(^2\) and there are an infinite number of administrative agency regulations. Secondly, there is the speed with which new material is being added to this already overgrown "corpus juris." Each year about 25,000 new opinions are published (nearly 700 cases per day) along with over 29,000 new statutes.\(^3\) The third reason is the overtaxing of the traditional indexing systems, and the increasing inability of these systems to meet the research needs of the lawyer whose clients are more and more demanding faster answers to their growing legal problems. Ancient legal concepts are being stretched to cover factual

---

* LL.B. Southern Methodist University; formerly Lecturer, Southern Methodist University School of Law; member of American Bar Association Special Committee on Electronic Data Retrieval; Vice President and Research Director, Southwestern Legal Foundation.

\(^1\) Address by Vincent Fiordalisi, Law Librarian, Rutgers University School of Law, to a Conference sponsored by the Electronic Data Retrieval Committee of the American Bar Association, August 29, 1960, in BNA, Proceedings of a Conference on Applications of Electronic Data Processing Systems to Legal Research 23 (1960).

\(^2\) Ibid.

\(^3\) Ibid.
situations undreamed of ten years ago, and new fields of law are constantly developing. Moreover, decisions of the United States Supreme Court in the past few years have opened the floodgates of litigation in areas which have lain dormant for over a century, e.g., civil rights, segregation, public education, and life insurance as interstate commerce, to name but a few.

The present indexing systems were developed in the early 1900's when law, like life, was slower and simpler, and when broad legal topics like “Negligence,” “Automobiles,” and “Master and Servant” (how archaic this sounds today!) were adequate terms with which to index the relatively few decisions then existing. A judge or lawyer then had time to read all the cases under the various subclassifications and select the ones most nearly in point.

Today, the situation is far different. Precedents have multiplied so rapidly that in Texas alone, for example, the topic “Negligence” in the Texas Digest lists over 6,000 separate digest entries which are organized under 144 Key Numbers (an average of about 42 items per number) and which occupy 341 pages. The topic “Automobiles” lists 6,400 entries and occupies 474 pages. Worst of all, of course, is “Appeal and Error” with over 40,000 separate entries and 1,247 Key Numbers. This colossal topic occupies four full volumes totaling 3,250 pages.

Further complicating the lawyer’s research job is the fact that in most conventional index-digests the headnotes state only the legal principles involved in the case. The factual background which makes the case relevant to a particular problem is either lacking or is so sketchy that the searcher must read the entire text merely to determine whether the case is in point.

Another cause of difficulty in the search of the law is the inflexibility inherent in the “hierarchical” indexes used in most current search books. Once a classification system has been established and numbers have been assigned to subtopics, the system tends to become stratified. Thus, it becomes very difficult to make changes with-

---

6 The statistics given are the results of an independent tabulation compiled by the author as of January 1, 1962. The results of the tabulation are on file with the Southwestern Legal Foundation, Dallas, Texas.
7 Ibid.
8 Ibid.
out seriously disarranging the other topics in the index, even though some rearrangements are obviously necessary to accommodate new fields of law or to provide for new emphases in existing fields. Such inflexibility places the index publishers in a serious dilemma. They must somehow fit all new decisions into existing index pigeonholes or else start a new topic—which means, in effect, scrapping their previous classifications.

A good example that illustrates this problem is the Decennial Digest, which is keyed to the National Reporter System. In the Fifth Decennial Digest (covering the years 1937 thru 1946) all social security cases are digested under Key Number 78.2 of the topic “Master and Servant,” where they occupy about 190 pages. However, in the Sixth Decennial Digest (covering the years 1947 thru 1956) “Social Security” became a major topic. It now occupies a single 1,800 page volume entitled “Social Security and Public Welfare” and has an entirely new organization of Key Numbers. As long as the research is confined to the one ten-year period covered by that particular Decennial Digest, little difficulty is encountered, but when a prior or a subsequent period is to be reviewed, the problem arises. Other examples of changes which the publisher has made to accommodate to the changing legal picture are the recent additions of “Aviation” and “Zoning” as major topics, and the substitutions of “Armed Services” for the old “Army and Navy,” “Labor Relations” for the old “Trade Unions,” and “Telecommunications” for “Telephones, Telegraph and Radio.” On the other hand, “Conflict of Laws” is still omitted, but “Common Scold,” “Livery Stable Keepers,” and “Embracery” (an intriguing subject) still remain as major topics, although they are merely lonesome holdovers from the distant past, with few or no precedents under them.

It must be remembered that the change of a major topic in a hierarchical index is not a matter of merely changing words, for many times the changes have repercussions down through a string of regional and state digests. It also requires a wholesale reorganization of the subtopics, not only within the major topic, but in the material left under the old topic.

Two other inherent characteristics of a hierarchical indexing system affect the research process adversely. First, the fact that each new decision must be boiled down to fit into a predetermined pigeonhole requires the digester either to leave out those portions of the case for which no pigeonhole exists, or to squeeze them, willy-nilly, into a preconceived mould. In either case serious distortion may result. Second, there are very practical limitations on the size of the index
volumes before they become unwieldy and on the size of the type before it becomes unreadable.

To summarize, the present day legal research picture consists of legal problems that are becoming more and more complex and yet in need of more rapid answers; an unwieldy accumulation of cases and statutes inherited from the past; a great yearly outpouring of new materials which must be added to the present accumulation; and indexing systems which are no longer precise enough to give access to pertinent precedents with sufficient speed or accuracy. As a consequence, legal research in important cases is unnecessarily slow and expensive. It results in delays in litigation, frustration for clients, and an inordinate expenditure of time and money on the part of lawyers. Thus, the time has arrived to look into the capabilities of modern scientific instruments, such as the electronic computers, to see if they can assume some of the research burden.

II. Meet an Electronic Computer

An electronic computer is a collection of four or more packing-case-sized boxes of various shapes which are wired together into a single functioning unit called a "system"—known colloquially as "hardware." The individual "boxes" making up the system are called "components," and each of the components performs distinct functions. These functions are described in the following paragraphs.9

Box A is the "input" component and is used to communicate with the machine. The standard method of communication is through the use of punched cards which are "key punched" on special machines by operators who work much like regular typists. Punched paper tape can be used as an alternative method of communication. On the cards or on tape, each letter of the alphabet or number has a characteristic pattern of holes. This pattern can be interpreted by the machine when the card or the tape is fed through an electrified reading head in Box A.

Box B is the central processor, or "brain," which controls and coordinates the activities of the other components by following an extremely detailed and precise list of instructions, called a "program." This program has been previously prepared and stored in the machine's memory. The chief feature of the central processor is its speed. It works at the speed of light, and its operations are timed in thousandths of a second.

Box C is primarily a storage unit for the material processed

---

9 The computer system is diagrammed in the appendix, figure 1.
by the machine. The two principal methods of storage are by magnetic tape and disc files. A “mag-tape” is a reel which holds about half a mile of metallic-coated plastic ribbon that can be magnetized in patterns of tiny dots. When the tape is passed at high speed across a “reading head,” the magnetized dot patterns are “read” by the machine. The tape must be read serially by beginning at one end and going through to the other, but it can be run at very high speeds. Each reel of tape holds more than one thousand cases in full text. Disc files are large sensitized metal plates stacked on a spindle with spaces between each disc for the “tone arm” or reading head to move. The spaces permit the tone arm to be set at any point on any of the discs, thus giving what is known as “random access” to anything stored on the discs.

As between discs and tapes, the tapes offer the more practical means for permanently storing large quantities of textual material such as cases and statutes. The advantages of tape are that it can be removed from the machine, kept on file, and, if necessary, “played” on other machines. Moreover, the information on one reel can be duplicated on another within a few minutes, thereby making it possible to furnish additional copies for use in other locations. Disc files, on the other hand, are permanently fixed to their spindle, are not portable, and are not suitable for permanent storage.

Box D, the high-speed line printer, is the most spectacular of all the components. Its function is to put the machine’s output into readable form, that is, into ordinary English text. Most printers now perform this function by printing at a rate of 600 lines per minute, which is fast enough for most purposes. In fact, a computer printer can finish in 3 minutes what it would take a competent stenographer all day to type. However, there are some limitations. First, there is only one font of type—all capital letters—and second, not all of the usual punctuation marks and conventional symbols are available.

It should be noted that Box B, the central processor, can be obtained in various “memory” capacities to meet the volume of work the machine will be required to do. Naturally, very large “memories” greatly increase the speed and power of the computer, but they cause a corresponding increase in costs. Similarly, from two to twenty units of Box C (using the tape units) may be connected together in an operating machine system to provide access to large quantities of stored material. Additional units or modules of disc files may also be wired into the system, if desired.

The machines used for punching the cards or paper tape are not

---

10 See figure 2 in the appendix for an illustration of the speed of Box D.
part of the “on line” hardware; that is, they are not connected with the computer system. However, they are vital parts and, of course, they must be available in order to allow communication with the computer.

Electronic computers were originally designed as counting machines, and their ability to add, subtract, multiply, and divide numbers is still their best known and most widely used talent. However, it is a mistake to think of a computer merely as a glorified desk calculator; for among its other functions, a computer can read and store alphabetic characters, carry out a “memorized” program of complex instructions, perform simple logic, and reproduce text in printed English. Two attributes in particular are of vital importance in searching authority by machine, and these same two clearly elevate the computer to a far higher plane than the calculator. They are: first, the ability to sort, or rearrange random words into a predetermined order; and second, the ability to match, or compare one word with another and subsequently to initiate action on the basis of the comparison.

III. How the Computer Indexes the Cases

Starting with a simple illustration, assume that the full texts of four state supreme court decisions have been punched into cards and stored on magnetic tape. These decisions shall be identified as documents 1, 2, 3, and 4. Acting under a program of instructions previously stored, the computer first breaks down document 1 into individual words and puts them in alphabetical order, listing each word only once regardless of how many times it appears in the document. The computer indicates by each word that the source is document 1. The machine then proceeds to sort documents 2, 3, and 4 in the same manner. It should be noted that this process would take hours, even days, if attempted manually; but the machine is capable of doing it in a few minutes. The result is that there are four word lists, one for each case, and that each word indicates its respective source-document number.

The next step is to merge the four lists into one, which the machine does by matching the words in each list. It then records each word and source-document number only once. For example, if the word “amount” appears in both documents 2 and 3, it would be listed only once with document numbers 2 and 3 following it. The machine has now created an index, but the kind of index never before used,
that is, one containing every word in all four cases together with the source references.\footnote{For an example of such a word index, see figure 3 in the appendix.}

By examining the combined word list, several things will be noted. First, it contains such words as "an," "any," and "are" and many others which are not really index words because of their inconsequential relationship to the contents of the respective documents. These words are called nonsignificant words and can well be eliminated from the index without any loss of information. Second, in the list there appear some incidental proper names of parties or witnesses such as "Ann" and "Antonio," together with a few figures representing dollar amounts, quantities, or dates. Only in rare instances are these words of any importance as index terms, and like the nonsignificant words, each can be eliminated from the final index which the machine will use in performing its automatic searching. The deletion of these nonsignificant words and incidental proper names from the index can be accomplished in several ways. One method in current use is to prepare a "common word list" in advance and instruct the machine to ignore any words appearing in this list. Another method is to keep these words in the vocabulary list but to give them a special identification number, such as zero, so that the machine can recognize them when preparing the condensed index to be used in conducting the search. Then any word bearing that number will be automatically excluded from the search index.

The third noteworthy feature about the word list is that the machine treats each different form of a word, including an occasional misspelling, as a separate word. For example, the machine considers "appeal," "appeals," and "appealed" as different words because of the difference in the detail of spelling or of the suffix. However, for indexing purposes, all of the words embody the same essential meaning and could be combined in the interest of shortening the list. The combination of these different forms of the same word is accomplished by the assignment of "root numbers." Each form of the word as it appears in the word list is assigned the same basic root number but different individual numbers. The machine then performs a numerical sort by root numbers and stores the result on a separate tape. The end product is a "Root Index File," where the root number stands for all forms of the root word and the source-document numbers stand for each root word which is collected.
There are six fairly distinct processes which make up an automated retrieval system. They are: (1) collecting the source material, that is, the "library" to be stored; (2) storing the material in machine-readable form and keeping it current; (3) indexing the material by machine; (4) searching the material; (5) reproducing the results of the search; and (6) disseminating the retrieved information. Since each of these processes has distinct problems, each is discussed separately below along with suggested solutions.

A. Collecting The Source Material

Several interrelated problems present themselves at the outset. First, there is the question of whether to store the documents, i.e., the reported cases, in condensed form or in full text. Considering the tremendous number of authorities to be handled and the bulky nature of the material, machine space and search time could undoubtedly be conserved by condensing the documents to their essentials before storing them. Furthermore, because publisher's headnotes and a few official digests are already available, assuming that copyright clearance is obtainable, these could be stored on a systematic basis without too much delay.

There are, however, many valid objections to storing pre-digested materials. Most existing digests were designed, as previously noted, to fit into a hierarchical indexing system. Consequently, many of the digest paragraphs become meaningless or even misleading when separated from their introductory catch words or preceding headnotes. Also, most digest paragraphs do not, and indeed cannot, purport to cover the whole case. The paragraphs represent what the digester felt were the essential elements in the decision—elements for which there happened to be pigeonholes in the publisher's index.15

Another possibility is to rewrite the cases in a condensed form. This approach, besides duplicating to some extent the work already performed by the publishers, would be an expensive and time-consuming task and would cause an undue postponement of the storage operation. In some states consideration is being given to the possibility of reviewing the older volumes of case reports with the goal of eliminating obsolete decisions and preparing official condensed rewrites of the remainder.16 Such efforts have, unfortunately, proved ineffective in the past and will probably continue to be so

---

16 Biunno, Expanding Case Law, November 1961 (unpublished paper on the New Jersey Reports located at 601 Broad Street, Newark, N.J.).
in the future. However, it would be most helpful to store any of the official condensations along with the entire text to supplement the decisions.

The third alternative is to store the cases in full text. This method has several disadvantages: (1) it is expensive to punch the text of the cases onto cards; (2) the full text occupies a considerable amount of space on the storage tape (in comparison to digested cases); and (3) the increased number of words require more complex machine indexing programs. On the other hand there are many important benefits. First, there is no need for human digesting, abstracting, or indexing of the cases before storage; second, each case may be key-punched verbatim from the official report; and third, key punch operators need not be trained in the law and need not exercise any discretion. Furthermore, the exact official text of the decision can be retrieved as output from the machine. Thus, there will be no need for maintaining expensive libraries at the computer center. This is obviously an important factor, especially where original documents are hard to locate, as in the case of most state and federal administrative regulations.

Other major advantages lie in the time-saving aspects. With full textual reprints available a researcher will not have to spend time looking for a case in the library but can read the full text as it is printed by the machine. Moreover, real progress is currently being made in developing optical scanning equipment which can read an entire printed page in seconds and convert the type images into magnetic spots on tape. When such machines become available (probably within less than two years) the input bottleneck will be solved and rapid storage of all the decisions will become possible. Another time-saving benefit is that it will be possible to use the actual words of the documents themselves as search terms; there will then be no need for creating special vocabularies of “descriptors,” “uniterms,” role indicators,” and other synthetic codes. Finally, and very important, a case will have to be stored only once. Thus, regardless of how many separate points of law are discussed, a case is amenable to machine search on any related subject and from any viewpoint after storage. Therefore, in view of the many advantages of storing cases in full text, this method has been adopted in most experiments.

Another problem encountered at the outset lies in determining the procedure to be followed in storing the cases. The alternatives are: (1) to select all the decisions in one area at a time and store them together (in which case the benefits of machine-aided research within
those areas would be obtained immediately); or (2) to store the
decisions serially regardless of their subject matter, just as they ap-
pear in the reporter volumes (in which case no complete search in
any field would be possible until all precedents had been completely
stored in the machine). Eventually, of course, all prior authoritative
decisions will have to be stored, but practical considerations of cost
and research dictate that for the present the first alternative be
adopted and that all the precedents be marshalled by fields.

For a pilot project the field of arbitration and award has been
selected. All of the appellate court decisions in this field, both state
and federal, from the courts of Arkansas, Louisiana, New Mexico,
Oklahoma, and Texas are being key-punched and stored in the
machine. Both labor and commercial arbitration cases have been
included. These cases number slightly under 200 and contain ap-
proximately 525,000 words of official text. Influential in the selection
of this topic was the fact that the cases contained a wide variety of
judicial prose styles, an extensive vocabulary of “fact words,” and a
large number of legal terms from many other fields of law, e.g.,
contracts, torts, labor law, real property, probate, and corporations.
Furthermore, there are enough decisions to provide a test library
and yet not so many as to prevent the inclusion of all the precedents
within the respective states. With the experience gained in machine
searching of multi-state problems in this one complete field of law,
it will very soon be possible to expand into larger and more difficult
areas, such as taxation and federal procedure, in response to the
needs (and financial support) of the profession and its clients.

Still another problem is the very practical one of gathering the
cases for key punching. Here the familiar procedures of conventional
research come into play, but with one important difference, namely,
that the searcher is here concerned only with decisions which deal
in some way with the area at hand, that is, he is not concerned with
the details. The primary reference sources used are the respective
state digests, supplemented by state law encyclopedias and annota-
tions appearing under the appropriate state statutes. In this type
of work the well-known and bothersome tendency to “branch out”
or to explore “fringe” areas provides assurance that all material on
the subject will be obtained. Armed with a deck of 5 x 8 inch white
indexing cards and a battery of pencils, the researcher notes the cita-
tions of the cases as he discovers them and does not worry about the
facts or the holding.

It must be pointed out that it is mandatory to establish at the
beginning a standard format for listing citations and to follow it
carefully. Citations are important because the title of the court, the state, and the year of the decision are index terms, or "keywords" in the search process. Consistency is necessary because the key punch operator will have to copy the citation from the 5 x 8 cards rather than from the reporter volumes. The citation format should be as follows: (1) names of the parties (using standard abbreviations for words such as "company," "corporation," "versus," etc.); (2) volume, reporter, and page numbers; (3) state name (abbreviated) and court; (4) year; and (5) history on appeal, if any. In keypunching under this format, the operator inserts a special symbol before the first letter of the state abbreviation and after the last figure in the year. This symbol (usually a diamond-shaped character) signals the machine that these words are from the citation and distinguishes them from state names and dates appearing in the text of the decision. By using this format it is possible to retrieve cases from one state alone, or from two or more specified states in any desired combination, and also to retrieve the cases by year of decision.

After all citations have been listed on index cards, the cards are then arranged in alphabetical order by names of the plaintiffs, and alphabetical lists are typed in several copies with ample space between the lines. These lists are then furnished to researchers who are instructed to discover and list as many additional cases as possible. The test for inclusion in the list is simple enough: Does the case mention (other than incidentally) or somehow pertain to the main subject? In doubtful instances, the case is to be included.

When the researcher has accumulated all the relevant cases, the list is then submitted to a noted authority for review. Finally, a master list is prepared to serve as a check on the key-punching progress and as a source for proofreading the citation cards for accuracy.

B. Storing The Cases In The Computer

As mentioned previously the text of the documents must be punched in a form which can be read by the machine. There are two methods of doing this. One method, developed for storing statutory material by John Horty, Director of the Health Law Center at the University of Pittsburgh, utilizes both paper tape and punch cards. In Horty's system the statutory sections to be utilized are first marked by hand (by the lawyer in charge of the

---

14 Senior law students with law review experience make excellent detectives.
15 A law professor or a private practitioner specializing in the field will usually examine the list in exchange for a copy.
They are then given to the operators who copy the text from the book word for word on a Friden Flexowriter, which produces a punched paper tape. The tape is run through an I.B.M. 047 papertape-to-card converter, which produces one punched card for each 75-character Flexowriter line of text at the rate of about 1,000 cards per hour. When a large deck of I.B.M. cards has been produced, the cards are run through an I.B.M. 407 printer, which furnishes a double-spaced print-out of the text. This is proofread against the original statute by two operators and corrections are marked directly on the printed sheet. The corrected copy is passed to editors who punch a correction card for each erroneous line using an I.B.M. 826 key punch machine connected to a typewriter. The correction cards are merged with the originals, and another printout is obtained and proofread. Any further corrections are then made. Finally, the cards are loaded into the card reader of the computer and "written" by the computer on magnetic tape. Horty reports that he has placed the entire text of the Pennsylvania statutes on tape as well as all the health statutes of thirteen other states at a cost of five to six cents per line.17

Another method of putting the material into machine-readable form has been tentatively adopted by the Southwestern Legal Foundation. In this method the full text of the court decisions, omitting headnotes and introductory material, is copied by keypunch operators directly on punch cards. The Foundation uses a printing punch that prints the text along the top of the card; this method allows the operator to note and correct the most obvious errors. The cards are then placed in the computer reader and "written" on a magnetic tape, called the "temporary source document tape," at the rate of 800 cards per minute. That machine also assigns each case a document number for identification. Next, the text is machine-sorted and matched against a master tape containing all previously stored words. New words, that is, words not appearing on the master word list, are punched on cards with their source-document numbers, one card per word. Simultaneously, the full text of the case is printed on paper for a cross-check against the original report. Correction cards are punched for all errors noted in the text print-out and are inserted in the original deck in place of the erroneous cards. The corrected deck is then used to write the Master Source Document

17 Horty, The Keywords in Combination Approach, M.U.L.L. (Modern Uses of Logic in Law) 58 (March 1962). M.U.L.L. is the quarterly newsletter of the American Bar Association's Special Committee on Electronic Data Retrieval, and is published in collaboration with Yale Law School. It is indispensable to anyone interested in logic and automated legal research.
Tape, which becomes the stored "library." New cases may be added to the "library" by the same process.

Next, cards which contain the new words (and their document numbers) are edited, and root numbers are assigned to each word or group of related words. It is at this point that "nonsignificant" words, such as "the," "and," and "of," are given a permanent root number of zero. This procedure eliminates them from the condensed index. The cards containing new words are then fed into the machine and are thereby added to the Master Word List. The new root numbers and new document numbers are placed on the Root Index File Tape.

It should be noted that one of the primary advantages of this system is that it enables much of the routine part of the storage operation to be performed by the machine. Manual efforts are therefore confined to those parts of the process which only a human can perform properly, namely, classifying the new words entering the system and correcting the text.

C. Indexing The Material By Machine

Indexing is the process of arranging and listing names, topics, and objects in order to facilitate the locating of individual items which are contained in the stored information. Some form of indexing must precede machine retrieval, since it is obvious that information cannot be retrieved until its location has first been ascertained. Present technology recognizes three pure types of index organization: hierarchical, subject-heading, and coordinate. Each system is analyzed below.

1. Hierarchical Indexes

In this type index each entry takes part of its significance from the preceding entries and contributes some necessary element to those following it. Such indexes generally progress from the most general terms at the beginning of a topic to the most specific at the end. Because they are so interrelated, it is quite difficult to insert new terms without seriously disrupting the whole scheme. Examples of this type of index are the Dewey Decimal System and the West Key Number Index-Digest.

2. Subject Heading Indexes

A subject heading index in its purest form consists of single words or phrases arranged in strict alphabetical order. There is no conceptual relationship between precedent or antecedent words. This is the most common type of index—in one form or another—and is found in the back of nearly all non-fiction books.
Subject heading indexes are seldom used in pure form, however, because of their length and the necessity for either extensive cross references or excessive repetition of terms and page numbers. Normally the type of index employed is a mixture of subject heading with a modified form of hierarchical. This system permits the use of general words as main headings and more specific words as subheadings. Sometimes there is even a further classification under the subheadings. The following is an example of a modified subject heading index:

**TENANTS IN COMMON [main heading]**
- Adverse possession [subheading]
  - ouster of cotenant not in possession [sub-subheading]
  - title obtained by one cotenant [sub-subheading]

As in all hierarchical indexes, the subheadings themselves do not give a complete portrayal of the contents; the preceding main headings are always an implied part of all subordinate headings and must be read with the subheadings to be understood. Also, it is evident from the above example that a person beginning his research with “Adverse Possession” rather than “Tenants in Common” may easily miss the reference altogether unless “Adverse Possession” has a subheading of “Tenants in Common.” Of course, a cross reference, e.g., “see Tenants in Common,” could be used to guide the reader.

The advantages of a pure subject heading index are its simplicity in locating a desired topic by the alphabetic arrangement and its flexibility in permitting insertion of new entries without disarranging the sequence. However, there are disadvantages. As the index grows, it acquires more of the hierarchical characteristics. This in turn leads to necessary but excessive repetition and the problems of where and when to enter new subheadings.

3. **Coordinate Indexes**

A coordinate index, as its name suggests, is one so arranged that the desired information may be found at the point where two or more known coordinates (index terms) “intersect.” A familiar (and distasteful) example of the coordinate index principle is the Standard Income Tax Table. To determine the tax one locates the appropriate marital status listed horizontally across the top and the amount of taxable income in the column at the left. The figure appearing where the two intersect is the amount of tax.

The coordinate index may be made to operate more mechanically and, if so, it is known as a “manipulative” index. One example of this system is the Termatrex, in which 8½ x 11 inch plastic sheets
represent selected index terms with tiny holes in the sheets identifying the source documents. When a research request is received, it is translated into keywords. The sheets representing these keywords are placed together on an illuminated table and covered by a piece of transparent plastic containing coordinate grids. A light beam passing through the matching holes in the sheets (indicating a document in which all the search terms coincide) falls on the grid coordinate. The coordinate code number is then used to locate the citations in an index. Some 4,500 decisions in the field of motor carrier law are currently being analyzed and manually indexed for a Termatrex system in Project Lawsearch, which is under the supervision of William H. B. Thomas, a Washington, D.C. attorney. The Termatrex apparatus itself will be marketed by its developer, Jonker Business Machines, Inc., of Gaithersburg, Maryland.

This same technique of passing a light beam through matching holes in a deck of keyword cards has been developed by I.B.M. technicians using standard 80-column punch cards. Automatic card sorting machines, known as collators, may also be used to locate pertinent documents. Under this procedure each card represents a document rather than a keyword, and the important features in the document are indicated by holes punched on the card. In using this method the pattern of holes characterizing the desired information is first determined. The cards are then run through the collator several times until only those cards with the prescribed pattern of holes remain. The citations may then be read directly from the selected cards. After the citations have been obtained, the cards must be returned to their normal places in the card file.

A type of index which contains traits of both the regular and the manipulative indices is the copyrighted Uniterm Index. This is an ingenious arrangement of hinged cards in book form which permits the searcher to combine several “Uniterms,” or artificial descriptive words, into one search. The system has been used to index United States chemical patents, and is available commercially from Information for Industry, Inc., of Washington, D.C.

In all coordinate index systems the retrieval operation is conducted in one of two ways, “look-up” or “search.” The “look-up” method may be visualized as having a keyword at the top of a page with its respective source document numbers listed in a column under it. Any document number appearing under all the keywords in a search request is considered pertinent. In the “search” method the

---

index is arranged with the document number on top and the keywords that describe its contents listed below it. Machine retrieval is performed by comparing the keywords in the search request with those on every card and by noting the document numbers whenever there is a match of the keywords. Under the "search" method the entire file of cards must be examined by the machine before a search is complete; whereas, in a "look-up" arrangement only those cards bearing the desired keywords need be fed through the machine.

D. Searching The Material

The search process is the very heart of any machine retrieval system; in fact, it can be said that if it is efficient, the entire system will be successful regardless of the capabilities of the input and output mechanisms. Conversely, if the search process is inefficient, the system will fail no matter how sophisticated the input and output operations or how large the memory capacity of the computer. The function of the search process is to furnish quickly, accurately, and economically all the pertinent authorities, and only the pertinent authorities, upon which a researcher may confidently rely with respect to the specific problems before him. The success or failure of the system is to be judged by how well it performs this function.

It is obvious that search objectives will differ according to the type of research project. For example, a legal scholar writing a treatise on the law of contributory negligence is interested primarily in seeing all preceding cases in which that general principle of law has been applied or discussed, and the factual backgrounds are of secondary importance. On the other hand, a practicing attorney defending his client against a damage suit for negligent collision at a blind street intersection is interested in locating all prior cases in his jurisdiction which involve cars colliding at blind intersections, whether they specifically discuss contributory negligence or not—that is, the facts are primary and the legal issues secondary, at least for the moment. If the attorney has access to all such cases, then he may completely decide whether to base his defense solely on the ground of contributory negligence or to combine it with other grounds discussed in the cases. A legislator faced with the problem of drafting an effective traffic control statute for his state will have another approach to the question, and an appellate judge deciding a case on appeal will have yet another. Furthermore, even the needs of the same legal researcher will differ from time to time even though the problem does not change. Thus, in designing a search system, it is important to realize that the pertinency of the re-
trieved materials to the problem at hand will vary considerably according to the individual researcher's needs.

1. The Elements of Pertinency

What makes a case pertinent as a precedent? A case is pertinent to a legal problem (1) when it deals with the same or analogous facts; (2) when it involves the same or substantially the same legal issues; (3) when it is by a court of the appropriate rank and geographical jurisdiction; and (4) when it constitutes the latest or one of the latest decisions on the point in question. These four characteristics which determine the pertinency of a case may be described as factual, legal, jurisdictional, and chronological elements. Each is discussed below.

a. The Factual and Legal Elements of Pertinency.—At first glance it would seem that fact situations in cases are so varied and involve so many details that it would be futile to attempt to use them as a basis for indexing. However, in reality facts in lawsuits, as set out in appellate decisions, tend to fall into patterns, and, strangely enough, they form at least as stable a basis for retrieving precedents as do the legal issues. One reason for this phenomena is that the facts and circumstances from which lawsuits arise have a striking similarity, regardless of the jurisdiction in which they occur. Collisions between automobiles, refusals to perform agreements, failure to pay debts, family disputes, and the other fertile causes of litigation follow substantially identical fact patterns whether they occur in Texas, New York, or Hawaii. Even more important, the courts usually describe these situations by certain “fact” words which are prevalent everywhere. On the other hand, not only does the substantive law differ from state to state, but the legal terms in which that law is couched vary among the different jurisdictions. The disparity is evident from a reading of the opinions in a particular area of appellate courts from several widely scattered states. Those portions of the opinions which state the facts display a substantial similarity in the words used, but the contrary is true of those portions which analyze the local law.

The significance of the widespread utilization of similar “fact” words is that with the computer, a single search can be made for both the factual and legal elements of pertinency by combining “fact” words and “law” words in the same search requests. This method will cause the machine to produce only those cases which have the desired combination of factual and legal issues.
b. The Jurisdictional and Chronological Elements of Pertinency.—
With respect to the jurisdictional element, as noted, it is important to be able to specify the particular geographical areas in which authority is desired, for obviously if a Texas lawsuit is the problem at hand, Texas precedents will be sought. Thus, there must be some method in machine search to select geographically pertinent cases.

The fourth element of pertinency is the date. Of course, the usual desire is for the latest cases, but it may also be desirable to restrict our search to the cases decided within a certain period. For instance, a statute may have been passed in 1940 and amended several times prior to its repeal in 1950. The researcher may desire the cases dealing with the statute. In this situation a search will be restricted to the decisions rendered since 1940. Similarly, a search may require the tracing of a chronological development of a legal doctrine through distinct historical periods.  

In summary, the electronic machine makes it possible to search simultaneously for all the elements of a case's pertinency, as compared to the conventional indexes where the search is limited to only one element at a time. A further advantage of machine search over contemporary methods is that the machine will perform an accurate factual search; whereas, with present indices, such is difficult, at best.

2. The Keywords-in-Logical-Combination Approach to Machine Searching

The development of a word index by the computor has previously been discussed. As noted, each judicial decision was divided into its component words and then consolidated into separate indices, one containing all the root words (including nonsignificant words) with their assigned root numbers, and the other containing the root numbers and their respective source-document numbers.

A typical legal problem shall now be proposed and its progress through the machine illustrated. Suppose a merchant has summarily discharged one of his employees for speeding in the company's delivery truck. The employee, through his union, contests the discharge and demands that his case be submitted to arbitration. An arbitration clause in the collective bargaining agreement provides that the company shall have the right to discharge an employee summarily "for good cause," but any other discharge shall be arbitrated. A memorandum discussing the law involved is desired.

The first step is to list all of the words which characterize the

---

19 A good example in Texas would be the history on the rights of married women.
problem, *i.e.*, words which may appear in the pertinent cases. The keywords for the above problem are then listed as follows:

arbitration
clause
employee
discharge
union
good cause
speeding
delivery
truck
collective
bargaining

This list of keywords constitutes the search request. (Synonyms are ignored in the interest of simplicity and will be discussed later.) The list is punched into the cards, which are then inserted into the card reader. From this point on, the search is completely automatic.

The keywords are first sorted by the machine into alphabetical order and are matched against the Master Word List to locate the correct root index numbers. At this stage the computer would print the products of its search as follows: (the numbers on the right are the root index numbers)

<table>
<thead>
<tr>
<th>keyword</th>
<th>index number</th>
</tr>
</thead>
<tbody>
<tr>
<td>agreement</td>
<td>0098</td>
</tr>
<tr>
<td>arbitration</td>
<td>0132</td>
</tr>
<tr>
<td>bargaining</td>
<td>0301</td>
</tr>
<tr>
<td>clause</td>
<td>0467</td>
</tr>
<tr>
<td>collective</td>
<td>0498</td>
</tr>
<tr>
<td>delivery</td>
<td>0986</td>
</tr>
<tr>
<td>discharge</td>
<td>1028</td>
</tr>
<tr>
<td>employee</td>
<td>1141</td>
</tr>
<tr>
<td>good cause</td>
<td>1874</td>
</tr>
<tr>
<td>speeding</td>
<td>3892</td>
</tr>
<tr>
<td>truck</td>
<td>4103</td>
</tr>
<tr>
<td>union</td>
<td>4631</td>
</tr>
</tbody>
</table>

These root numbers obtained from the Master Word List are passed against the Root Index File, and the machine notes and lists the source-document numbers for each root index number. It then compares the various lists of source-document numbers. When this process is completed, the printed results, with the pertinent document numbers underlined, will appear as follows: (The left hand
column contains the root numbers, each representing a root word and its various forms; the numbers to the right represent the source document numbers.)

By comparing the source document numbers with each other the machine determines that numbers 20 and 79 meet all the requirements of the search request, i.e., those documents contain all of the specified words. To round out the search, the machine orders Box C, which contains the source document text tape, to locate those two documents.

E. Reproducing The Results Of The Search

Upon finishing the search, the machine commands Box D, the printer, to print documents 20 and 79 from the tape. This order is fulfilled at the rate of ten lines per second. The sheets are then removed from the printer, folded once down the center to form two pages, arranged in proper page sequence, and stapled in the left margin to form a pamphlet. This process completes the search.

It is possible to run as many as fifty separate searches at the same time merely by numbering each search request and identifying each keyword by attaching an appropriate search number. The machine will go through the process just described with all the search words and then separate them by search numbers before the documents are printed.

F. Disseminating The Retrieved Information

The final stage in electronic computer retrieval concerns the problem of getting the search output into the hands of the user. In many instances, of course, the user will be present when the output comes from the machine and can at that moment read it. On the basis of
the first results, additional searches can then be made. Also, the output may be mailed or hand delivered to a researcher's office if it is at some distance from the computer center. However, overshadowing both of these conventional methods is the recent development of data transmission devices which permit direct intercommunication between outlying "stations" and a central computer.

Typical of the new dissemination systems is Western Union's Broadband Switching System. A subscriber is provided with a "voice/data" instrument which is a little longer than a telephone and with which he may contact the desired computer center. Using the voice circuit he may first talk to the computer center, and then by punching a button he will be connected with the machine for high speed data transmission. "Data will be fed through the sending machine. It will be received and digested by a miles-away computer, and the answers will be fed right back." Provision is now being made to transmit facsimiles of retrieved material consisting of photocopies of court decisions, agency regulations, or lists of pertinent citations.

It may be noted that several large corporations now utilize an intra-company network of communications between smaller computer installations at various branches and the major computer, which is normally located at the home office. Individual law offices could be connected in a similar manner to computers located at major legal information centers or large law libraries. Such centers would make it possible to establish on a cooperative basis depositories of scarce legal materials. An individual firm or lawyer could have access to these materials when needed without the unnecessary expense of purchasing and housing them.

V. The Synonym Problem

One of the traditional glories of the English language is its richness in synonyms. But whatever blessing this abundance may be to literary men, it is a bane to lawyers, scientists, and all others who must communicate complex ideas simply and precisely. As a classic example of synonyms in the law, consider the following combinations of words which may be used to express the simple idea of an "oral agreement."

---

21 Ibid.
22 Collins Radio Company, for example, presently has a large computer at its Cedar Rapids, Iowa office connected with a smaller computer at its Dallas office.
Each of these twelve words is different (parol and parole are two different words to the machine) yet all are mutually interchangeable. Thus, there are thirty-six possible ways of saying essentially the same thing.

From the machine retrieval standpoint the problem posed by synonyms is the possible failure to retrieve a pertinent case because the search request did not contain the same word or words which were used by the court in writing the opinion. Although this problem exists in any index, it is, in theory, a special hazard in computer search. The reason is that the search process is not visible to the researcher, and he is not able to correct errors or follow new leads as he progresses.

On the other hand there are several factors which reduce the possibilities of missing a case through failure to use the proper search term. The first factor is the tendency of courts to repeat the same concepts in different words at various places in the opinion or to quote statements from other decisions which include different words. This repetition is a natural thing, for redundancy is as tiresome to a writer as to a reader. Accordingly, in the vast majority of opinions there are several “word handles” for each concept. Naturally, these “handles” greatly increase the probability that a “hit” will be achieved in the search process.

Another factor is that the speed of output permits a researcher to add new terms to his request to make the search more precise. Conversely, he may enlarge the scope of the question by subtracting terms if the results of the first search so warrant.

A third factor is the lawyer’s familiarity with the terms and phraseology of legal literature. Normally, a law-trained researcher has acquired a feel for the traditional modes of expression and vocabulary of the law from years of reading cases and writing briefs. Thus, he knows in advance the several varieties of factual and legal terms which are likely to appear in the decisions pertinent to his problem. This special knowledge, incidentally, is a factor often overlooked by lay critics of the keyword search approach; unfortunately, they tend to visualize the typical searcher as a layman adrift without a compass in a sea of legal verbiage.
The fourth factor, closely related to the preceding one, is that the keywords are taken from the law cases themselves, and no outside sources or specially-coined words are necessary. Indeed, the keyword approach has been adopted because it eliminates the need for learning a new vocabulary, and because, rather surprisingly, it permits a substantial reduction in the number of terms which the machine (and the researcher) must handle. (The details of this reduction and other statistical data on the number of words in the judicial vocabulary are discussed below.)

The fifth and most important factor of all is that the searcher will have a printed “thesaurus,” or list of synonyms and related words obtained from the text of the cases, to aid him in preparing a search request. The thesaurus is so arranged that all synonyms and near-synonyms are grouped together or cross-referenced, thus insuring that important words will not be overlooked and, conversely, that no “false” keywords, that is, words not found in the cases, will be used as search words.

A. The Function Of A Thesaurus

The value of a thesaurus as an aid to machine retrieval is now generally recognized, although there is some disagreement among retrievalists as to what its function should be. Actually the role of the thesaurus is largely determined by the type of search system employed. In systems which use synthetic keywords as search terms, the thesaurus must take on, to some extent, the character of a dictionary. In those systems it must carefully define any restrictions on or expansions of word meanings. In numerically coded systems, the thesaurus lists the English words followed by the code numbers, plus a reverse-index listing the code numbers first and then their English equivalents. In some systems the thesaurus lists not only synonyms and near-synonyms, but also generic, related, and associated terms. A few retrieval systems use no thesaurus at all but have the researcher pick some search words by conjecture and then run them through the machine to establish “profiles” of associated words. A rather complex formula is then applied to the profile to determine an “association factor,” and the terms most often associated with the original terms are used in locating the most pertinent documents.

B. Preparing The Thesaurus

The preparation of a thesaurus for computer research in case law is essentially a manual operation with a substantial assist from the machine. The suggestion has been made that the process might be
rendered completely automatic by storing the contents of Roget's Thesaurus, Webster's Dictionary of Synonyms, and Rodale's Synonym Finder on tape and using them as word sources. However, there are several valid objections to this proposal. Aside from the time and substantial expense of key punching these lengthy volumes, it is extremely doubtful that a thesaurus compiled from such sources would be precise enough to serve the needs of a lawyer. Basically, a legal researcher requires a tool for getting access to specific items in a specialized library of stored documents which are characterized by their own peculiar vocabulary and modes of expression. Thus, the more sensible approach, it seems, is to prepare a thesaurus in the terms of this specialized library rather than in the literary terms of reference books such as Webster's Dictionary. This does not mean, of course, that purely "literary" terms should be excluded from a legal thesaurus; however, it does mean that if they appear, they do so because they are found in the text of the stored library in some context appropriate to the case.

The argument for storing the contents of law dictionaries and Words and Phrases is on stronger ground, for these words do embody legal applications. On the other hand, those who suggest machine storing of these excellent references overlook the fact that the problem is not to find a general source of legal terms (the cases in the stored library provide enough of these) nor even to find a guide to legal synonyms generally. The specific problem at present is to provide law-trained inquirers with a guide to the keywords and their synonyms which are found in a particular library of reported decisions or statutes. Admittedly, however, a thesaurus which is derived from a multitude of decisions does offer many possible benefits, and the idea is receiving further study. At the present time it is too early to predict the best source. Consequently, experience and development will have to be the guides.

VI. Future Machine Developments

Judging from the progress in machine technology to date, the information retrieval system of three years from now will look quite different from the four-box mechanism previously described. By then there will be special purpose machines expressly designed to store and retrieve information; whereas the present general purpose machines are built to accomplish a variety of data processing and computing chores. To be more specific, a new computer, tentatively named the "I.R.-1," will feature a number of improvements. Input will be accomplished by feeding the printed pages of law
books between rotating rollers, which are strikingly reminiscent of clothes wringers. The rollers will position the printed page in front of an electronic scanning “eye,” which through a complex maze of circuits will convert the image of the printed characters into standard patterns of magnetized dots on a reel of computer tape at a rate of approximately 2,000 characters per second. No punched cards or punched paper tape will be utilized in the input process. With such equipment the problem of storing the vast accumulation of legal literature will essentially be solved. For processing, the storage capacity of machine memories will be enlarged, and new materials will permit the construction of memories large enough to hold tremendous quantities of text material in suspended animation. All documents will be subject to instant retrieval on command.

Output for permanent use will be letter- or legal-size hard copy photographed from stored microfilm cards, reels, or scrolls. When permanent copies are not needed, as where the searcher desires to “browse” through a stored collection of documents or to check a quotation, an image screen will be available on which the desired pages can be projected.

VII. Preliminary Statistics on the Words in Court Decisions

Although the greatest interest undoubtedly centers around the ability of a computer to retrieve pertinent authorities, there are some interesting by-products of the retrieval process which may prove to be of importance in applying proven scientific statistical principles to certain aspects of the law. One of these by-products is a capability for studying in depth the vocabulary of judicial opinions and statutes. From such studies it may be possible to find ways to shorten future court decisions, to eliminate ambiguities in meaning and, perhaps, to condense existing law libraries without loss of precision.

Preliminary testing was done on a collection of twenty cases in the area of arbitration and award from the appellate courts of Arkansas, Oklahoma, and Texas. The cases were chosen at random without regard to length or contents. The full text of each decision was transcribed onto punched cards, which were then “read” onto magnetic tape. Utilizing machine programs specially written for the purpose, the words of the text were counted, combined, and sorted into alphabetical order. Each word was listed only once, regardless of the frequency with which it appeared in the twenty cases. Headnotes and editorial catchwords were omitted.
The twenty cases contained a total of 30,229 words with the average case being 1,515 words long. When each word was listed only once, however, the resulting list (called the master word list) was merely 2,771 words long (less than ten per cent of all of the words stored)—thus indicating a very high rate of repetition. The master word list contained a high proportion of nonsignificant words, i.e., words not significant as index terms. These totaled 535, and were comprised of 298 different general words without index value, such as “the,” “and,” “therefore,” and “witnesseth”; 183 personal names, e.g., Smith, Tom, and Ferguson; 24 of the 26 letters in the alphabet, “x” and “y” alone being absent; 17 names or abbreviations for months and days of the week; and 13 different abbreviations for regional reporter systems. The master word list also showed every variant of a root word (for example “allege,” “alleged,” “alleges,” and “alleging”) as a separate word. When all the nonsignificant words were eliminated, and all root word variants were reduced to one example, the resulting keywords totaled only 1,367, or less than five per cent of the total words stored.

The significance of the foregoing statistics is that there need be stored a much smaller vocabulary for case law search than had been anticipated. Also, the decrease in words stored will produce resulting benefits in speed and accuracy of search.

As more and more cases are stored the figures will become more accurate. It is too early to say definitely but present indications are that there will not be a substantial increase in the number of keywords. The first goal is to get the entire body of case law in one field stored in computer-searchable form. With the experience gained in this field, steps can be taken into larger and more complex areas according to the needs of the profession and the resources available.
APPENDIX

Fig. 1. A diagram of a computer retrieval system.
MORGAN V. TEEL 234 P 200 JOKLA 1925) REH DEN
JARMAN, C. THE PETITION OF THE PLAINTIFF, J.C. TEEL.
AS AMENDED ALLEGES: THAT ON OR ABOUT JANUARY 1, 1920, THE
PLAINTIFF RENTED FROM THE DEFENDANTS JOHN H. MORGAN AND ANNA
A. MORGAN CERTAIN LANDS FOR PASTURE AND AGRICULTURAL PURPOSES
FOR THE YEARS 1920 AND 1921, THAT THE LAND IN QUESTION WAS
OWNED BY THE DEFENDANT ANNA A. MORGAN., THAT THE DEFENDANTS
WERE TO FURNISH CERTAIN LIVE STOCK WHICH WAS TO BE FED OUT OF
THE FEED STUFF GROWN ON THE PREMISES, AND THE PLAINTIFF WAS
TO LOOK AFTER AND CARE FOR SAID LIVE STOCK, AND THE
PLAINTIFF AND THE DEFENDANTS WERE TO SHARE EQUALLY IN THE
PROCEEDS DERIVED FROM THE SALE OF SAID LIVE STOCK., THAT THE
DEFENDANTS PURCHASED CERTAIN HOGS WHICH WERE HANDLED IN THE
SAME MANNER, AND THE PLAINTIFF WAS TO RECEIVE TWO-THIRDS OF
THE PROCEEDS DERIVED FROM THE SALE OF SAID HOGS, AND THE
DEFENDANTS WERE TO RECEIVE ONE-THIRD OF SAID PROCEEDS.,
THAT, DURING JULY OF 1920, THE DEFENDANTS LEASED SAID
PREMISES FOR OIL AND GAS PURPOSES, AND THE LESSEES AND THEIR
ASSIGNS WENT UPON SAID RIGHTS OF THE PLAINTIFF THEREIN, AND,
DURING THE EARLY PART OF 1921, DIFFERENCES AROSE BETWEEN THE
PLAINTIFF AND THE DEFENDANTS RELATIVE TO THE USE AND
OCCUPATION OF SAID PREMISES, RESULTING IN ANNA A. MORGAN, BY
AND THROUGH HER AGENT, JOHN H. MORGAN, ENTERING INTO A WRITTEN
AGREEMENT WITH THE PLAINTIFF TO ARBITRATE THEIR DIFFERENT
CLAIMS, AND IN KEEPING THEREWITH THREE ARBITRATORS WERE
SELECTED TO SETTLE SAID DIFFERENCES, RESULTING IN THEIR
AWARDING TO THE PLAINTIFF THE SUM OF $1,650 AS THE VALUE OF
THE USE OF THE LEASED PREMISES FOR THE REMAINDER OF THE YEAR
1921, AND UPON CONDITION THAT THE PLAINTIFF SHOULD SURRENDER
POSSESSION THEREOF, AND, ON THE SAME DATE AND IN A SEPARATE
AGREEMENT, THE DEFENDANT ANNA A. MORGAN, BY AND THROUGH HER
DUTY AUTHORIZED AGENT, JOHN H. MORGAN, AGREED WITH THE
PLAINTIFF THAT THE LIVE STOCK ON THE PREMISES AT THE TIME
SHOULD BE ADVERTISED AND SOLD AT PUBLIC SALE AND THE PROCEEDS
DIVIDED ACCORDING TO THEIR SEVERAL INTERESTS THEREIN.. THAT
THE AWARD OF THE ARBITRATORS AND THIS SPECIAL AGREEMENT TO
SELL THE LIVE STOCK WERE MADE ON FEBRUARY 24, 1921., THAT,
IN PUSUANCE OF SAID AGREEMENT, BILLS WERE PRINTED,
ADVERTISING THE SALE OF SAID STOCK TO BE HELD ON MARCH 11,
1921, BUT, ON MARCH 2, 1921, THE DEFENDANT ANNA A. MORGAN
NOTIFIED THE PLAINTIFF THAT SHE WOULD NOT BE BOUND BY THE
ARBITRATION AND THE PLAINTIFF TOOK NO FURTHER STEPS TO SELL
ANY OF THE LIVE STOCK.
THE PLAINTIFF PRAYS FOR JUDGEMENT SUSTAINING THE AWARD
MADE BY THE ARBITRATORS AND FOR THE SALE OF THE LIVE STOCK
AND FOR A DIVISION OF THE PROCEEDS, DERIVED THEREFROM,
ACCORDING TO THE INTERESTS OF THE PARTIES ABOVE SET OUT.
THE DEFENDANTS ADMIT THAT THE LIVE STOCK WAS TO BE
FURNISHED TO THE PLAINTIFF ON THE BASIS SET OUT IN THE
AMENDED PETITION OF THE PLAINTIFF AND ALLEGE THAT SAID STOCK
WAS TO BE FED OUT OF THE FEED STUFF GROWN ON SAID PREMISES,
BUT THAT, DURING THE YEAR OF 1920, THE DEFENDANTS FURNISHED
FEED FOR SAID STOCK IN THE SUM OF $355 WHICH WAS EXCLUSIVE OF
AND IN ADDITION TO THAT PRODUCED ON THE PREMISES DURING THE
YEAR 1920., THAT THE PLAINTIFF DID NOT PREPARE THE GROUND
FOR CROPS NOR DO ANYTHING TOWARDS PLANTING AND RAISING CROPS
ON SAID LAND DURING THE YEAR 1921, BUT WAS LETTING THE SAME
LIE OUT, TO THE DAMAGE OF THE DEFENDANTS IN THE SUM OF $800.,
THAT THE PLAINTIFF HAD NOT CARED FOR AND LOOKED AFTER THE
STOCK IN THE PROPER MANNER AND AS AGREED TO, AND THE
DEFENDANTS PRAY FOR JUDGMENT AGAINST THE PLAINTIFF FOR THE
DAMAGES SUFFERED AND THAT THE SALE OF SAID STOCK BE
PROCEEDED WITH.

UPON THE ISSUES THUS JOINED, JUDGMENT WAS RENDERED FOR
THE PLAINTIFF, SUSTAINING THE AWARD, IN THE SUM OF $1,612.50,
AND GIVING THE DEFENDANTS THE POSSESSION OF SAID PREMISES
UPON THE PAYMENT OF SAID AWARD, FROM WHICH JUDGMENT THE
DEFENDANTS HAVE APPEALED.

THE DEFENDANTS URGE TWO PROPOSITIONS FOR THE REVERSAL
OF THIS CAUSE. FIRST, THAT THE ARBITRATION WAS NOT BINDING
UPON THE DEFENDANT ANNA A. MORGAN, FOR THE REASON THAT JOHN
H. MORGAN HAD NO AUTHORITY TO BIND HER IN ANY AGREEMENT WITH
RESPECT TO ARBITRATION., SECOND, THAT THE PLAINTIFF WAS NOT
ENTITLED TO THE AWARD OF $1,612.50 AND ALSO THE POSSESSION OF
THE LAND FOR THE YEAR 1921.

THE DEFENDANTS ARE NOT IN POSITION TO URGE THEIR FIRST
PROPOSITION HEREIN, FOR THE REASON THAT THE MATTERS WHICH
THEY NOW SEEK TO HAVE THE COURT TO CONSIDER FOR THE PURPOSE
OF AVOIDING THE AWARD MADE BY THE BOARD OF ARBITRATORS WAS
NOT PLEADED BY THEM., AND IT IS PLAIN FROM AN EXAMINATION OF
THE RECORD AND PLEADINGS IN THE CASE THAT THIS
PROPOSITION IS AN AFTERTHOUGHT. THE CASE WAS NOT TRIED ON
THAT THEORY. ARBITRATION IS A SUBMISSION OF DISPUTED MATTERS
TO SELECTED PERSONS FOR THEIR DETERMINATION AND THE
SUBSTITUTION OF THEIR DECISION OR AWARD FOR A JUDGMENT BY
COURTS, AND THIS METHOD OF SETTLING CONTROVERSIES IS
RECOGNIZED BY THE COMMON LAW, AND, AS WE HAVE NO STATUTE
ON THE SUBJECT, THE COMMON LAW PREVAILS. DEAL V. THOMPSON,
51 OKL. 256, 151 P. 856. THE AWARD OR DECISION OF THE
ARBITRATORS HAS THE SAME FORCE AND EFFECT AS THE JUDGMENT
OF A COURT, AND DISPUTED MATTERS WHICH ARE THUS ADJUSTED
CANNOT AFTERWARDS BE RETRIED IN AN ACTION AT LAW, UNLESS
THE PLEADING STATE FACTS SUFFICIENT TO AVOID THE AWARD.
SCRIVNER V. MCCLELLAND, 67 OKL. 51, 168 P. 415. THE FACTS
PLEADED BY THE DEFENDANTS ARE INSUFFICIENT TO AVOID THE AWARD
OF THE ARBITRATORS, AND IN FACT NO ATTEMPT WAS MADE BY THE
DEFENDANTS IN THEIR PLEADINGS TO AVOID SAID AWARD AND
THEREFORE THE FIRST PROPOSITION SUBMITTED HEREIN CANNOT BE
CONSIDERED.

AS TO THE SECOND PROPOSITION, THAT THE PLAINTIFF IS NOT
ENTITLED TO THE AWARD OF $1,612.50 AND ALSO THE POSSESSION
OF THE LAND FOR 1921, IT IS SUFFICIENT TO SAY THAT THE
JUDGMENT OF THE TRIAL COURT DOES NOT GO TO THAT EXTENT. IT
GIVES TO THE PLAINTIFF JUDGMENT FOR THE AMOUNT OF THE AWARD,
AND TO THE DEFENDANTS THE POSSESSION OF THE PREMISES FOR 1921.

FOR THE REASONS HEREBEFORE ASSIGNED, THE JUDGMENT OF
THE TRIAL COURT IS AFFIRMED.

Fig. 2. A computer-printed case in full-text. The printing took only nine
seconds. The two parentheses in the citation are used as symbols to identify
the state and date as keywords.
<table>
<thead>
<tr>
<th>CASE</th>
<th>WORD</th>
<th>COUNT</th>
<th>CASE</th>
<th>WORD</th>
<th>COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAPE 1</td>
<td>PRINT</td>
<td>RECORD</td>
<td>TAPE 1</td>
<td>PRINT</td>
<td>RECORD</td>
</tr>
<tr>
<td>ALLEGATIONS</td>
<td>2</td>
<td>57</td>
<td>APPEAL</td>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>ALLEGED</td>
<td>1</td>
<td>58</td>
<td>APPEALED</td>
<td>1</td>
<td>86</td>
</tr>
<tr>
<td>ALLOW</td>
<td>3</td>
<td>59</td>
<td>APPEALS</td>
<td>3</td>
<td>87</td>
</tr>
<tr>
<td>ALLOWANCE</td>
<td>1</td>
<td>60</td>
<td>APPEAR</td>
<td>1</td>
<td>88</td>
</tr>
<tr>
<td>ALLOWED</td>
<td>3</td>
<td>61</td>
<td>APPEARS</td>
<td>3</td>
<td>89</td>
</tr>
<tr>
<td>ALLOWING</td>
<td>3</td>
<td>62</td>
<td>APPPELLANT</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>ALONG</td>
<td>2</td>
<td>63</td>
<td>APPELLEE</td>
<td>2</td>
<td>91</td>
</tr>
<tr>
<td>ALREADY</td>
<td>1</td>
<td>64</td>
<td>APPELLEES</td>
<td>2</td>
<td>92</td>
</tr>
<tr>
<td>ALSO</td>
<td>1</td>
<td>65</td>
<td>APPLICABLE</td>
<td>1</td>
<td>93</td>
</tr>
<tr>
<td>ALTERNATIVE</td>
<td>2</td>
<td>66</td>
<td>APPLY</td>
<td>1</td>
<td>94</td>
</tr>
<tr>
<td>AM</td>
<td>4</td>
<td>67</td>
<td>APPOINTED</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>AMER</td>
<td>3</td>
<td>68</td>
<td>APPOINTMENT</td>
<td>2</td>
<td>96</td>
</tr>
<tr>
<td>AMICABLE</td>
<td>2</td>
<td>69</td>
<td>APPROVAL</td>
<td>3</td>
<td>97</td>
</tr>
<tr>
<td>AMONG</td>
<td>3</td>
<td>70</td>
<td>APPROVE</td>
<td>3</td>
<td>98</td>
</tr>
<tr>
<td>AMOUNT</td>
<td>2</td>
<td>71</td>
<td>APPROVED</td>
<td>3</td>
<td>99</td>
</tr>
<tr>
<td>AMOUNTED</td>
<td>1</td>
<td>72</td>
<td>APRIL</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>AMOUNTS</td>
<td>3</td>
<td>73</td>
<td>ARBITERS</td>
<td>4</td>
<td>101</td>
</tr>
<tr>
<td>AMPLE</td>
<td>2</td>
<td>74</td>
<td>ARBITRARIES</td>
<td>3</td>
<td>102</td>
</tr>
<tr>
<td>AN</td>
<td>1</td>
<td>75</td>
<td>ARBITRATE</td>
<td>1</td>
<td>103</td>
</tr>
<tr>
<td>AND</td>
<td>1</td>
<td>76</td>
<td>ARBITRATION</td>
<td>1</td>
<td>104</td>
</tr>
<tr>
<td>ANGLETON</td>
<td>3</td>
<td>77</td>
<td>ARBITRATOR</td>
<td>2</td>
<td>105</td>
</tr>
<tr>
<td>ANN</td>
<td>3</td>
<td>78</td>
<td>ARBITRATORS</td>
<td>1</td>
<td>106</td>
</tr>
<tr>
<td>ANOTHER</td>
<td>4</td>
<td>79</td>
<td>ARCHITECT</td>
<td>3</td>
<td>107</td>
</tr>
<tr>
<td>ANSWER</td>
<td>1</td>
<td>80</td>
<td>ARE</td>
<td>1</td>
<td>108</td>
</tr>
<tr>
<td>ANSWERS</td>
<td>2</td>
<td>81</td>
<td>ARISING</td>
<td>3</td>
<td>109</td>
</tr>
<tr>
<td>ANTONIO</td>
<td>3</td>
<td>82</td>
<td>ARK</td>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>ANY</td>
<td>1</td>
<td>83</td>
<td>AROSE</td>
<td>2</td>
<td>111</td>
</tr>
<tr>
<td>APP</td>
<td>3</td>
<td>84</td>
<td>ARRIVE</td>
<td>1</td>
<td>112</td>
</tr>
</tbody>
</table>

Fig. 3. A portion of the combined word list for four cases, prepared by machine. The first column of figures gives the number of the document in which the word first appeared. The second column gives the word count.