Vanderbilt Law Review

Volume 19
Issue 3 *Issue 3 - June 1966*

Article 11

6-1966

Legal Research--Computer Retrieval of Statutory Law and Decisional Law

David T. Moody

Legal Research—Computer Retrieval of Statutory Law and Decisional Law

I. The Problem of Legal Research

Legal research presently involves a considerable amount of any lawyer's time and efforts largely because it is a slow and tedious process. Searching for a pertinent legal point can prove to be time-consuming and often fruitless. Moreover, it is here that chance plays one of its largest roles in the law.[1] An important legal point may exist, yet the researcher may fail to find it although he exercises a great degree of diligence. All lawyers must recognize this problem and the fact that it is becoming more acute with the passage of time. Something needs to be done to facilitate legal research in order to preserve the efficient development and administration of law by the courts, legislatures and, more importantly, by the lawyers in their day-to-day practice.

An analysis of the problems of legal research is appropriate at this point. The bulk of the present law is largely responsible for the difficulties encountered in legal research. The accumulation of statutes and cases carries with it the necessity that the lawyer cover much more material than in the past. Thus, the conscientious lawyer of today must devote much more time to researching a legal point than did his predecessors. To give an indication of the problem, it has been determined that in the United States, there are one and one-half million statutory sections and two and one-third million reported cases in existence today.[2] Each year over 29,000 new statutes are passed and about 25,000 new opinions are published.[3] This increase in the volume of existing law makes the lawyer's research problem even more acute.

This constant increase in legal material creates a further problem of indexing. Present research books use hierarchical indexes which begin with a broad legal principle followed by narrower sub-headings. "Once a classification has been established and numbers have been

1. "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 seeking. This, in turn, undesirably increases the element of chance in the trial of cases and the planning of business affairs." Dickerson, *The Electronic Searching of Law,* 47 A.B.A.J. 902 (1961).

2. Wilson, *Computer Retrieval of Case Law,* 16 Sw. L.J. 409 (1962), quoting Vincent Fiordalisi, Law Librarian, Rutgers University School of Law.

3. Fiordalisi, *Panel Discussion: Progress and Problems in the Application of Electronic Data Processing Systems to Legal Research,* 1960 Modern Uses of Logic in Law 174, 175-76 [hereinafter cited as M.U.L.L.].

assigned to subtopics, the system tends to become stratified."[4] When a new statute is passed or a new decision reported, the indexer must either fit the statute or decision into the existing hierarchy or make changes in the index to accommodate the new statute or decision. While some rearrangement is obviously necessary to accommodate new fields of law or to provide for new emphasis in existing fields, it becomes very difficult to make many alterations without seriously disarranging other topics in the index.[5] If the indexer chooses to fit the new statute or decision into the existing index, he may either have to eliminate certain portions for which no classification exists or to squeeze them into existing classifications. Both of these procedures may result in serious distortions of the statute or case. Thus, the inflexibility of present indexing systems may seriously hamper the researcher's efforts.[6]

Present indexing systems often provide no assistance for the lawyer who is searching for a particular factual situation. Most indexes refer to legal principles only. The factual background which makes the case relevant to a particular problem is either omitted or is so abbreviated that the lawyer must read the entire text to determine the relevance of a case to his factual situation.

Therefore, the present bulk of the law, its rate of increase, and the inflexibility and inadequacy of present indexing systems may present a formidable barrier between the lawyer and the answer to his legal problem—a barrier which may be difficult, if not impossible, to hurdle. The computer, however, may provide a solution to these problems.

II. THE COMPUTER

There are many types of computer systems. These systems vary in size, complexity, speed, cost, and application. However, all computer systems involve at least three basic considerations: the source data or *input* entering the system; the orderly, planned *processing* within the system; and the end result or *output* from the system. Processing is carried out according to a pre-established sequence of instructions that are followed automatically by the computer. The plan of processing is always of human origin. By calculation, sorting, analysis, or other operations, the computer arrives at a result that may be either used for further processing, or recorded as reports or files of data.

4. Wilson, *supra* note 2, at 410.
5. See *Id.* at 411 for a good example of this problem.
6. See generally Davis, *Automatic Data Processing and the Judge Advocate General's Corps*, 1964 MILITARY L. REV. 117, 122-23.

Communication with a computer system requires that data be reduced to a set of symbols that can be read and interpreted by the computer. These symbols differ from those commonly used by people, because the information to be represented must conform to the design and operation of the machine. The choice of these symbols and their meaning is a matter of convention on the part of the designers. The important fact is that information can be represented by symbols, which become a language for the communication between people and machines.

Information to be used with the computer systems can be recorded on several media, some of which are: punched cards, paper tape, magnetic tape, magnetic ink characters, and optically recognizable characters. Data are represented on the punched card by the presence or absence of small rectangular holes in specific locations of the card. In a similar manner, small circular holes along a paper tape represent data. On magnetic tape, the symbols are small magnetized areas, called spots or bits, arranged in specific patterns. Magnetic ink characters are printed on paper, and the shape of the characters and the magnetic properties of the ink permit the printed data to be read by both man and machine. The shape of the optical characters, together with the contrast with the background paper, permits optical characters to be read both by the machine and also by people.

All computer systems can be divided into four types of functional units: input devices, output devices, storage, and central processing unit. The computer system requires, as a necessary part of its information-handling ability, features that can enter data into the system and record data from the system. These functions are performed by input-output devices. An input device is a machine designed to sense or read information from one of the recording media. In the reading process, recorded data are converted to electronic form; the data then can be used by the machine to perform computer operations. An output device is a machine that receives information from the computer system and records the information on either punched cards, paper tape, magnetic tape, or as printed information on paper.

All data must be placed in storage before they can be processed by the computer. Information is read into storage by an input unit and is then available for internal processing. Each location, position, or section of storage is numbered so that the stored data can be readily located by the computer as needed. The computer may rearrange data in storage by sorting or combining different types of information received from a number of input units. The computer may also take the original data from storage, calculate new information, and place the result back in storage. The size or capacity of storage determines

the amount of information that can be held within the system at any one time. In some computers, storage capacity is measured in millions of digits or characters. The capacity and design of storage affect the method in which the data are handled by the system.

The central processing unit is the controlling center of the entire computer system. It can be divided into two parts: the arithmetic-logical unit and the control section. The arithmetic-logical unit performs such operations as addition, subtraction, multiplication, division, shifting, transferring, comparing, and storing. It also has logic ability—the ability to test various conditions encountered during processing and to take the action called for by the results. The control section directs and coordinates the entire computer system as a single multipurpose machine. Its functions involve controlling the input-output units and the arithmetic-logical operation of the central processing unit, and transferring data to and from storage. This section directs the system according to the procedure originated by its human operator.

Each computer system is designed to perform only a specific number and certain types of operations. It is directed to perform each operation by an instruction, defining a basic operation to be performed and identifying the data, device, or mechanism needed to carry out the operation. The entire series of instructions required to complete a given procedure is known as a program. Thus, a complex problem must first be reduced to a series of basic machine operations before it can be solved. Each of these operations is coded as an instruction in a form that can be interpreted by the computer and is placed in the main storage unit as a stored program.

There are several characteristics of a computer which make it highly desirable for dealing accurately with a large source of material. First, the computer is fast. The central processing unit works at the speed of light, and its operations are timed in millionths of a second. The cards reader input component can read cards at a rate of up to 1000 cards per minute. Magnetic tape can be read at rates of up to 340,000 characters per second. The printer output component can print at rates up to 1,100 lines per minute.[7]

A second desirable characteristic possessed by the computer is accuracy. The computer will manipulate data with 100 per cent accuracy. However, the absolute accuracy of the computer's output is limited by the accuracy of the data fed into the computer and the capability of the program to accurately handle these data. Therefore, if the input data and programs are accurate and adequate, the output will be accurate, thus removing the possibility of human error in manipulation of the data.

7. These figures are based on maximum rates of the IBM System/360 computer.

If the computer can be used to search the massive accumulation of legal material, perhaps the present bulk of the law would not pose such a difficult problem in legal research. Also, the accuracy of the computer would insure that the possibility of oversight in legal research would be minimized. Therefore, it appears desirable that the computer be utilized to cope with the problem of legal research. In recent years, significant steps have been taken in this direction. An examination of computer retrieval of statutory law and decisional law follows.

III. Searching Statutory Law by Computer

The application of computers as a tool for performing research of statutory law has achieved a high degree of success due to the efforts of John F. Horty and his staff at the University of Pittsburgh Health Law Center. The "Key Words in Combination" system developed by Mr. Horty is a complete computer-searching system for a body of law.[8] The Pittsburg organization originally applied this method to the Pennsylvania statutes.[9] The following is a description of the procedure for developing a computer searching system for statutory law.[10]

The entire text of the statute sections are placed on punched cards. The punched statutory text is then fed into the computer which in turn places the statute sections on magnetic tape. Each statutory section is put onto magnetic tape word for word, exactly as enacted. Each section of the statute is treated as a document and as the section is placed on the magnetic tape, it is given a document number. At the same time, the computer creates on another magnetic tape an alphabetical list of every word found in the statutes followed by the exact location of each occurrence of the word in the statutes. The location identifies the document number (which refers to a statute

8. See Horty, *The "Key Words in Combination" Approach*, 62M M.U.L.L. 54 (1962).

9. Mr. Horty's organization has now applied this method to the entire statutes of New York, New Jersey and Ohio, and the statutes which deal with health in eight other states. The New Jersey Constitution, court rules and rules of evidence have been included as well as the ordinances of two municipalities: Middletown, New Jersey, and Pittsburgh, Pennsylvania. Mr. Horty's "Key Words in Combination" approach has also been applied to case law which presently includes the Pennsylvania Superior Court decisions and the Pennsylvania Supreme Court decisions. In addition, the opinions of the Attorney General of Pennsylvania dealing with education are included. The United States Air Force Accounting and Finance Center in Denver, Colorado, employed Mr. Horty's organization to apply this method in Project LITE—Legal Information Thru Electronics. The bodies of law included are the entire United States Code and all volumes of the Comptroller General Decisions, as well as the unpublished decisions. In addition, the LITE system includes the Armed Services Procurement Regulations.

10. This method may be utilized for case law also.

section), the sentence number within the statute section, and the word number within the sentence.[11] Certain very common words are omitted from the alphabetical list of statute words, thus words such as "the," "a," "therefore," are omitted.[12] This alphabetical list is the basic tool for searching statutes. After the full-text tapes and the vocabulary-location (index) tapes have been created,[13] the computer search of the statutes may begin.

Once the researcher has determined his problem, he selects words and their synonyms which might appear in a statute dealing with the problem. In short, the researcher selects words which make the statute sections he seeks unique as compared to the rest of the statute sections on tape. As an aid to prevent exclusion of relevant words and phrases, the searcher should frequently refer to a list of words actually used in the statutes or a specially-constructed thesaurus based upon words used in the statutes. Upon completion of a relevant word list, the researcher gives the list to a computer operator who prepares the list for search by computer. The words are fed into the computer and, at the same time, the computer finds each word on the vocabulary tapes and notes all the document locations of each word. After all words have been processed in this manner, the computer then "searches" the statute-text tapes and prints out the citation or full text (whichever is desired) of each statute section (document) in which the relevant words appear. In order to make the search more precise, the researcher may specify that each document contain more than one relevant word, or he may list groups of words and specify that each document contain a word from each group. In this manner, the researcher can make his inquiry as broad or as narrow as he wishes.[14]

---

11. The entry for the word 'tax' might appear thusly on a printout . . .

Tax	897.009.7	4281.013.5
	4282.016.4	4921.004.11

The word 'tax' appears in document 897, sentence 9, as word 7, it appears again in document 4281, in sentence 13 as word 5." Horty, *supra* note 8, at 58.

12. In the Pennsylvania project, 112 words were omitted from the alphabetical list. These words were, for the most part, articles, pronouns, conjunctions, etc.

13. The statutes in Pennsylvania, in full text, consist of 32,323 statutory sections (documents) including 650,808 lines. These materials comprise five reels of text tape and five reels of vocabulary tape. Excluding 112 common words which are not used, the total Pennsylvania vocabulary consists of 23,406 different words in 353,172 sentences. There are 6,230,529 total words in the statutes of which 2,259,407 are common words and 3,971,122 are those words found in the vocabulary.

14. The following is an example given by Mr. Horty. "Thus if he [the researcher] wished to search for all the Pennsylvania statutes dealing with illegitimate children, he might put into one class the words 'baby,' 'child,' 'foundling,' 'infant,' 'juvenile,' 'minor,' 'orphan,' etc., along with their various forms. Thus requiring that one of these words, at least, appear in a document for it to be considered relevant. Another class could be established containing the words 'father,' 'mother,' 'parent,' 'unwed,' 'unmarried,' 'legitimate,' etc.

The short period of time required by the computer to perform a search is remarkable. To search through the vocabulary tapes for the list of relevant documents (statute sections) and to print out citations to these documents requires between twenty-five and forty minutes on the IBM 7070 computer.[15] Also, there is no limit to the number of searches the computer is able to process simultaneously. The machine limitation is that no more than 500 words may be searched for during any one pass through the vocabulary tape.[16] Thus, several searches may be processed simultaneously to obtain maximum efficiency from the system.

In addition to speed, the accuracy of this system has been tested and proven favorable.

> To provide an indication of the efficiency of the retrieval system, six computer scarches . . . were also searched manually. . . . Results of the two types of searches were compared. For purposes of comparative analysis an 'A,' 'B,' & 'C' grading scale was used. An 'A' statute was deemed to be one which a researcher would wish to read in solving his research problems whether it ultimately proved to be of use or not. A 'B' statute was deemed one which the searcher would wish to read if certain facts, not clearly stated in the inquiry, were actually present. A 'C' statute was one deemed not relevant to the inquiry.
>
> Computer searches of these six inquires yielded a total of 177 'A' statutes, 9 'B' statutes and 143 'C' statutes. In contrast, manual searches of the six inquiries, conducted by members of the law school faculty, produced 72 'A' statutes, 9 'B' statutes and 32 'C' statutes. The number of 'C' statutes retrieved by the manual searchers is a minimum figure since it is known that the manual searchers scanned and disregarded without notation many statutory provisions. The manual search produced 2 'A' statutes, and 2 'B' statutes which the machine searcher did not retrieve.[17]

"To specify to the machine the relationship which must exist between the words in context, a certain operator is utilized. One such operator is the word 'or,' which is used within each class above to tell the machine that either 'baby' or 'child' or 'foundling,' etc., must appear in the document for it to be considered relevant. When it is desired to tie two classes or two words together, the operators 'D,' 'S,' or 'W' may be used. If, as was done in the search above, it is decided that the statutory section contain at least one word from the first class of words and at least one word from the second, the operator 'D' is used to indicate that at least one word in each class must appear in a relevant document. Similarly, if a tighter relationship is desired, the operator 'S' would be used to indicate that representatives from each of the classes must appear in the same sentence.

"In a search involving illegitimate children, in addition to those documents containing representatives of the two classes stated above it may be desired that certain documents be considered relevant if a certain single word appeared therein, such as "illegitimate,' 'bastard,' 'parentage,' 'relative,' etc. If the document containing the phrase 'born out of wedlock' is sought, the operator 'W+3+3' is used. This operator requires that the word 'wedlock' appear in the same sentence, no more than three words after 'born.' " Horty, *supra* note 8, at 60.

15. *Ibid.*
16. *Ibid.*
17. *Ibid.*

The striking aspect of the Pittsburgh project is that the computer is given a completely unindexed body of literature, and it creates its own index based upon the actual language of the statutes. Once the researcher has provided the words relevant to his problem, the computer can search the entire body of statutes with much greater accuracy and speed than the human researcher. Also, the option is not limtied to statutory law. It is apparent that this system may also be effectively applied to administrative regulations, executive orders, and the like, since these have the characteristics of statutes such as brevity and clarity. "It is quickly apparent that the Pittsburgh system is the first to bring to bear on legal research problems computer capabilities other than speed. In the area of statutory law, where the system has been largely utilized, its effectiveness has been sharply demonstrated."[18]

IV. SEARCHING DECISIONAL LAW BY COMPUTER

There have been several approaches to retrieval of decisional law by computer. Three significant methods will be dealt with here.

A. *Root Index System*

An approach to research in the field of decisional law, similar to Mr. Horty's in statutory law, was developed by Robert Wilson, Research Director of the Southwestern Legal Foundation.[19] Mr. Wilson's system of automated searching of full natural text of case law will be described briefly.

The full text of the court decisions, omitting headnotes and introductory material, of a particular field of law or of a particular geographical area[20] is key-punched onto punch cards. The punched cards are then fed into the computer which assigns each case a document number and places the document number and case on magnetic tape. Next, the computer creates an alphabetical vocabulary list of all words found in the cases followed by the document numbers of all the cases in which the word appears. The computer records

18. Eldridge & Dennis, *The Computer as a Tool for Legal Research*, 28 LAW & CONTEMP. PROB. 78, 88 (1963).

19. For a detailed description of Mr. Wilson's system see Wilson, *supra* note 2, at 409.

20. For the Southwestern project, the field of arbitration and award was selected. All of the appellate decisions in this field, both state and federal, from the courts of Arkansas, Louisiana, New Mexico, Oklahoma, and Texas were included in the project. There were slightly less than 200 cases which contained a total of approximately 525,000 words of official text. *Id.* at 419.

each word and document number only once.[21] It should be noted that the computer treats each different word form including a mispelling, as a separate word. For example, the computer considers "appeal," "appeals," and "appealed" as different words because of the difference in the detail of the spelling; however, for indexing purposes, all of the words embody the same essential meaning and could be combined to shorten the list. The combination of these different forms of the same word is accomplished by the assignment of "root numbers." This is where the human element enters the indexing process. Each form of the word, as it appears in the computer word list is assigned the same basic root number. The computer then creates two magnetic tape files, one which contains each word found in the cases followed by its root index number, and a second which contains each root number followed by the document numbers of all the words which the root number represents. Thus, there are three magnetic tape files: one with the full text of all cases, a second containing each word found in the cases followed by its root index number, and a third containing each root index number followed by the document numbers (case numbers on the full text tape) where the root number represents all forms of the root word and the document numbers represent each case in which that root word is found. This system is now ready to be utilized for research.

After analysis of his problem, the researcher makes a list of all the words which might characterize the problem. The list is then punched into punch cards, and the punched cards are fed into the computer. The computer sorts the list in alphabetical order and then matches the words against the tape file containing each word of the cases followed by the root index number. The computer then punches out each word submitted by the researcher followed by its root index number. These root index numbers are then matched by the computer against the magnetic tape file containing each root number followed by the document numbers of the words represented by the root number. The computer then obtains all relevant document numbers. At the end of the search, the computer matches the retrieved document numbers against the magnetic tape file containing the document numbers followed by the full text of the cases, and the full text of each relevant case is printed by the computer. The search may be broad or narrow in scope depending on the decision of the researcher. If the researcher specifies many words that must be found in each case to make it relevant, then the scope of the search may be very narrow.

21. For example, if the word "amount" appears several times in both documents 2 and 3, it would be listed only once with the document number 2 and 3 following it. *Id.* at 414.

In the Southwestern project, it is possible to run as many as fifty separate searches at the same time.[22]

Presently, the only major drawback to this system is the difficulty that would be involved in placing all of the cases on magnetic tapes—even those in a single state. This problem was not encountered in the statutory retrieval system. The presence of the problem here stems from the difference in length of statute sections as compared with cases which tend to be many times greater in length. However, future technological development in the computer industry may well overcome this obstacle and pave the way for widespread use of the Southwestern system or the Horty system in the area of case law. The system developed by Mr. Horty seems preferable since his indexing process is fully automated, whereas the Southwestern system injects the human element into the indexing procedure.

B. *Point of Law System*

The "Point of Law" method of researching case law was developed by the late Robert T. Morgan, assistant professor of business law at Oklahoma State University.[23]

The first step of the point of law approach involves the analysis of each case for the pertinent issues or "points of law" that are considered in the decision. When a particular point is determined, it is given a code number. These code numbers, therefore, represent legal concepts. This analysis is performed by humans. One magnetic tape file contains a document number for each case, followed by the citation of each case along with the headnotes, or full text, or both. Another magnetic tape file contains each point of law followed by the document numbers (case numbers) which deal with the point of law. The researcher first analyzes his problem in terms of legal concepts. He then looks up in the alphabetical listing of points of law and submits the corresponding code numbers to be searched by the computer. The computer searches the magnetic tape file containing each point of law followed by the document numbers and obtains all relevant document numbers. Next, the computer searches the magnetic tape file containing each document number followed by the case represented by the document number. The computer then prints all relevant cases by citation, citation and headnotes, or citation, headnotes, and full text. This system adds nothing to conventional indexing except speed, and therefore the problems involved with current

22. *Id.* at 428.

23. For a more complete description of Mr. Morgan's system see Morgan, *The "Point of Law" Approach*, 62M M.U.L.L. 44 (1962).

indexing methods are present here. " 'The Point of Law' approach may be characterized as an automated and vastly accelerated West Key Number tape system."[24] The only difference from present manual methods of research is that several points of law can be searched at one time.

C. *ABF-IBM Project*

The American Bar Foundation and the International Business Machines Corporation jointly conducted a project to explore application of computers to legal research.[25] The full texts of approximately 5,000 cases, taken chronologically from the Northeastern Reporter, were key-punched into punch cards, and the punched cards were fed into a computer which created a magnetic tape file containing the full text of each case preceded by an identifying document number. The cases were not restricted to any one field of law. The following briefly describes the method utilized.

A thesaurus called an "index-word space" is first constructed. The computer "reads" approximately one-half or 2,500 cases from the tapes, and obtains word frequency counts. The frequency of each unique word that appears in the cases is used to determine the skewness of its distribution in the file. It is theorized that words which characterize cases (documents) with respect to the other cases in the tape file will have a skewed or uneven distribution while words which do not characterize cases will have an even distribution throughout the file. For example, the word "the" would be found in about the same proportion in each case and would therefore be evenly distributed throughout the cases. This would be a non-informing word. On the other hand, the word "slander" would be found a number of times in cases dealing with slander and very little, if at all, in cases not dealing with the subject. Therefore the word "slander" would be distributed unevenly throughout the cases which would result in a skewed distribution. This would be an informing word. After obtaining the distribution of each word in the file, some skewness cutoff point is determined, and each word is placed into one of two categories—Type I (non-informing words) or Type II (informing words). A dictionary on magnetic tape of the non-informing words is then constructed by the computer.

A list of "informing words" is prepared for each of the 2,500 cases. For each informing word, an "association factor" is calculated for every other informing word with which it appears in any one case by computing the probability that Word A would appear this close

24. Eldridge & Dennis, *supra* note 18, at 86.
25. See *ibid.* for a detailed description of the proposed system.

to Word B this number of times over the entire file if the informing words were distributed at random. Small probabilities indicate a high degree of association and high probabilities a lesser degree. These numbers are used to estimate the distances between the words in the index-word space.[26]

From this information, an index-word space is constructed in which every word is placed at the correct distance from every other word in the system with which it exhibits association.

> The result of this operation can be visualized schematically as a sort of grid in which every word can be placed in its appropriate position by assigning it a set of coordinates . . . in such a way that the set of coordinates for any individual word consists of a string of six digits.[27]

An informing word dictionary on tape is then prepared where each word is followed by its six-digit set of coordinates in the index-word space. The tape dictionaries of the informing words and the non-informing words constitute the computer thesaurus. The informing word dictionary on tape is the index-word space in which words that appear geometrically close together are closely associated in some way, and words that appear far apart are not closely associated. Therefore, concepts can be represented by volumes carved out in the index-word space. "[T]hese volumes may be said to correspond to the descriptors or index terms that might have been attached by humans, without their arbitrariness of classification."[28]

The remaining cases on the tapes (other than the 2,500 already dealt with) are indexed by the computer "reading" each case from tape and using the non-informing word tape dictionary to discard words and the informing word tape dictionary to prepare an analysis of word frequencies related to index-word space. Instead of specific words as indexing terms, concepts are selected as volumes in index-word space.[29] The document numbers (case numbers) are then filed in document space, where the grid matches the index-word space one-for-one. This document file is on magnetic tape.

Searching the cases for pertinent legal or factual points is performed in the following manner. The researcher submits the question narratively. This question is key-punched into cards and fed into the computer. The computer is then to index the question in the same

26. *Id.* at 97.
27. *Ibid.*
28. *Id.* at 98.
29. "A rough physical analogy to this process would be to toss pennies at the previously mentioned grid so that, for every Type II (informing) word in the source document, a penny lands at its proper slot on the grid. Where the pennies heap up in a pile, you have a concept. The circumference of the base of the pile defines the scope of the concept." *Ibid.*

manner it indexed the cases. The non-informing word tapes are used to discard words in the question, and the informing word tapes are used to determine word frequency counts in the question, thus determining the concept volumes that represent the question. Once the concept volumes have been obtained, these volumes are searched in document space (on the document space tape) for the relevant document numbers. After obtaining the relevant document numbers, the computer searches the tapes containing the document numbers followed by the full texts of the case, and the relevant cases are printed out by the computer. As this system builds up by the addition on new cases, it will have a tendency to become more accurate because the analysis of each word is based upon a greater number of cases.[30]

A highly desirable characteristic of the case law retrieval systems utilized by the Horty system, the Southwestern project, and the American Bar Foundation-IBM propect is that factual situations as well as legal issues are indexed by the computer and may be quickly searched. This could not be done effectively by present research methods. Fact searches are necessary to increase search effectiveness and pertinency of result. A fact search may be an end in itself. Fact searches may yield cases which reveal legal issues not previously recognized by the searcher. Furthermore, a fact search can be used to narrow a large yield of cases from previous searches.

V. Conclusion

Powerful as the computer may seem, the lawyer cannot be replaced. The lawyer must obtain the facts and characterize the problem. At this point the computer may perform the research based upon the lawyer's characterization. After obtaining the relevant material from the computer search, the lawyer must analyze and evaluate the material and apply the law based upon his evaluation.

However, the potential impact of computer science upon the law is great. The speed of computers makes the bulk of the law relatively insignificant. The potential flexibility of indexing by computers may render present indexing systems obsolete. The lawyer of tomorrow

30. "A conceptual analogy can be drawn between the proposed system and a learning human. The system develops a sense of the meaning of words by 'reading' extensively. It cannot know anything about a word that it has seen only once, but as it experiences the word repeatedly in different contexts, it begins to 'catch on' to its meaning (sometimes erroneously). The more it reads, and the better the quality of the material it reads, the better informed will be the system. It will become an expert in its field, although it may prove quite ignorant when first turned loose in another's field! When this manner of ignorance becomes a significant lack, the system can be advised to read in another area to close the gap." *Id.* at 99.

may notice a significant reduction in his research time, thus increasing his ability to provide services to his clients. With the removal of the human element from the actual researching and considering the thoroughness of computers, the quality of the lawyer's services, as well as the quantity, may increase. If opposing lawyers have "all" the law, the possibility of ill-founded court decisions would more than likely decrease, and the possibility of compromise between parties would increase, thereby relieving some of the caseload burden upon congested court dockets.

Presently, lawyers may submit requests for all statutes or cases dealing with a particular legal or factual topic. Provided the statutes or cases of the particular jurisdiction have been prepared for computer research, the lawyer may receive the total text of all relevant statutes of cases dealing with the topic for a nominal fee.[31] In the light of the demonstrated accuracy and speed of the system, the demand for such service should increase as the legal profession becomes more familiar with its use.

In the future, law firms may subscribe to legal information centers located in major cities for these services. Thus, each member of the firm could receive all legal material pertinent to his problem within a short period of time. With such quick access to legal materials, a law firm may find it no longer necessary to incur the expenses of purchasing and maintaining them in a private library. Thus, computer science may have a tremendous impact upon the law in the very near future.

DAVID THOMAS MOODY

31. Mr. Horty's organization, located at the Pittsburgh Health Law Center in Pittsburgh, Pennsylvania, is currently rendering this service.