

Computer-Assisted Legal Research: Some Fundamental Concepts

OLD AND NEW PROBLEMS

The legal profession relies heavily on the effective accumulation, storage, and retrieval of information. Of the various kinds of information used by lawyers, the most vital is the large body of primary materials which constitute the written law — the statutes, regulations, and judicial decisions. The term "legal research" usually denotes a selective process of locating, interpreting, and applying elements from the primary sources. Effective methods of accessing the primary materials are therefore indispensable.

Traditional access methods have relied on indexes, abridgments, and encyclopedias, for the simple reason that these were the only tools available. Utilization of these access tools depends on a hypothetical "meeting of minds": to locate desired information, the researcher must look under the right heading or classification, and to do this he must choose the same terminology as the indexer. This process can involve more guessing than choosing. One is repeatedly faced with the task of deciphering classification schemes, of determining by trial and error how the indexer categorized particular concepts or fact situations. This initial phase of legal research can often be time-consuming, and in some cases research attempts may begin and end at this phase.

Traditional access methods also rely heavily on the availability of primary sources and secondary finding aids. In terms of volume and expense, these materials present an increasingly serious problem. Even if expense is not considered, growing litigious and statutory/regulatory activity increases the need for more efficient retrieval methods.

Most would probably agree that traditional access tools have improved in the last decade. The emergence of more subject reporting series and specialized loose-leaf reporters bring together cases, legislation and relevant materials bearing on selected areas of the law. The recent adaptation of key-word index schemes patterned on U.S. models has significantly facilitated access to some Canadian case reports and statutes. These refinements are welcome, but they are piecemeal. The question is whether larger, more detailed printed indices constitute the sole solution to present problems, or whether new and radically different access methods are desirable. This is a relatively recent question given substance by a relatively recent product of technology: the computer.

COMPUTER-ASSISTED LEGAL RESEARCH

Several recent attempts to improve access to legal materials have involved computers. Adaptation and development of the new technology to serve the unique requirements of legal research is a complex and ongoing activity, but our present concerns are relatively simple:

- 1) How can the computer assist in practical legal research?
- 2) How does the computer work?

These questions are really two aspects of a single question, and they suggest the difficulty which lawyers often experience when confronted with the new method, because the first question can be answered only after the second question has been answered. In order to appreciate the value (and limitations) of the new method, it is desirable to acquire a basic understanding of how the computer works.

HOW THE COMPUTER WORKS

Q/L Systems Ltd. (formerly QUIC/LAW) maintains a central computer located in Kingston, Ontario. Over a period of more than ten years, various legal materials have been entered into the computer, and permanently stored in its memory. These materials are grouped together in compartments called data bases. Examples of legal data bases are: the Statutes of New Brunswick; the Atlantic Provinces Reports (headnotes); the Dominion Law Reports (headnotes).¹ As time passes, more data bases are created (the Statutes of Ontario are presently in the process of being entered), and existing data bases are expanded by the addition of headnotes of recent cases or statutory amendments. Data bases vary in the time-periods they cover. For example, the DLR data base contains at the time of writing all the headnotes to cases in the Dominion Law Reports reported between September, 1955 and January 7, 1982; the RSC data base is composed of the office consolidation of Federal acts in force at December, 1981, prepared by the Department of Justice.

Once legal materials have been entered into the computer's memory, computer programming makes selective retrieval possible. The selective retrieval method presently in use is a comparison method. The researcher communicates with the computer *via* a terminal which is connected to the computer by a special telephone service called Datapac. The terminal resembles a typewriter. A word or combination of words are typed into the computer *via* the terminal. The computer will then search for this word or combination of words in each of the documents in the chosen data base. For example, in the DLR data base, the computer will search for the word or words in every DLR headnote back to September, 1955.

¹See complete list of legal data bases to date in Appendix A.

This search process takes about ten seconds, after which the computer will respond by activating the terminal and advising you how many times your word or word-combination appears in the data-base you are searching, and in how many documents it appears. You can then instruct the computer to print out the documents on your terminal.² You may instruct the computer to order the output of documents in one of three ways. You may specify:

- 1) that the document in which your requested term(s) appear most frequently be displayed first; next most frequently second; etc.
- 2) that the documents appear in chronological order;
- 3) that the documents appear in reverse-chronological order.³

Combining Words

If you enter two or more words joined by "&", the computer will retrieve only documents containing *all* of those words.

If you enter two or more words joined by a space, the computer will retrieve all documents containing any one or more of those words.

The "but not" command (%) is used to exclude documents containing the words which follow it.

Example: marihuana marijuana & sentence % importing.
All documents containing either of the two spellings of the word "marihuana" and also the word "sentence", *except* those which contain the word "importing", will be retrieved. Documents which do not contain one of the two spellings of the word "marihuana" will not be retrieved. Documents which do not contain the word "sentence" will not be retrieved. Documents which *do* contain the word "importing" will *not* be retrieved.

A phrase may be searched by simply enclosing it in quotation marks:

Example: "real market value".

You can combine words and phrases:

Example: "real market value" & appraisal.

²On terminals equipped with a CRT (cathode ray tube), you can examine the documents, one by one on the screen, and selectively print those which you want on an attached printer.

³Further refinements of this ranking system are possible. See QL Systems Limited, *QL/Search User's Manual*, (QL Systems Limited, July 1, 1981), 150.

To search words *or* phrases, or phrases *or* words, you must use a special technique.⁴

The following observations apply to the basic functions listed above.⁵

The Computer Searches for Words, Not Concepts:

The computer does not recognize meaning in a word or phrase — it simply treats a word as a series of letters, evenly spaced. Thus, a search for the word “contracts” will not retrieve documents which contain the word “contract” but not “contracts”. A fortiori, a search for “contract” will not retrieve “agreement”; “marijuana” will not retrieve “cannabis”.

Since documents on the computer have not been grouped and classified under subject headings as is the case with encyclopedias and abridgments, it is necessary to adopt a new research approach. Before performing a computer search, one must ask, “What words are likely to appear in a headnote attached to the type of case which I am looking for?” This can be a difficult habit to cultivate, since one is inclined by force of habit to apply the conceptual approach to computer searching. The conceptual approach will work only to the extent that the headnotes have been captioned by the familiar heavy-print classification terms which are usually (but not always) included after the name of the case.

Obviously, the element of uncertainty in traditional manual research is not removed by computerized research. An informed guess is involved in looking for words which are likely to appear in a headnote. However, important distinctions must be noted. First, the computer offers access to each and every significant word contained in the headnotes. The resulting increase in alternative access points permits the lawyer to apply his knowledge of judicial and legislative language to the full text of the headnotes, eliminating the possibility of distortion created by an intervening indexer. Second, the possible access points change not only in number, but in kind: search words which are descriptive of *fact* situations may be used. The computer thus offers a new way of retrieving relevant cases. For example, to locate cases which consider the question of liability in a situation where a retail business listing has been omitted from the yellow pages of a telephone directory, one may choose to search for the phrase “yellow pages”, on the assumption that any headnote dealing with the issue would include that phrase.

While the multiple access terms offered by the computer broaden the possibility of retrieving relevant headnotes, they also create problems.

⁴A recent change in the system has made this possible by the creation of sets. *Ibid.*, 115.6.

⁵Several other important functions, such as the ability to search word-stems (truncation), are omitted from this fundamental discussion. See QL/Search User's Manual for detailed descriptions of all Q/L functions. *Ibid.*

In examining these problems, one should keep in mind a pivotal assumption of computerized research, which is that those documents which contain the term(s) which the researcher has chosen will, in fact, be relevant.

Relevancy and Noise

A "relevant" document is a document that you are looking for. "Irrelevant" documents may deal with your issue but not in a way that is useful, as when circumstances render the facts distinguishable. "Noise" denotes totally irrelevant documents, as when a search for the word "battery" meant as a tortious act retrieves headnotes containing references to car batteries.

A number of irrelevant documents, and some noise, is to be expected in most searches. As a general rule, the more comprehensive your search is, the higher will be the percentage of irrelevant documents and noise. It is sometimes possible to eliminate irrelevancy and noise completely by a very narrow and restrictive formulation of the search terms. But the price paid for total relevancy is often high, because by formulating the search terms narrowly, relevant headnotes can easily be missed. For this reason, total relevancy is often a indication that the terms should be reformulated. The presence of some irrelevant documents is often reassuring, since it may indicate that the search was sufficiently broad. For example, one may search "assault" or "assault & battery", but decline to search "battery" alone, because of anticipated noise. The result will probably be no noise, but those headnotes which used the word "battery" but not the word "assault" will be missed.

The process of question-formulation is most often an attempt to strike a balance or compromise between comprehensiveness as the desired goal, and irrelevancy/noise as a necessary inconvenience. In complex searches this balance will vary according to the purpose of the searcher. Ideally, the lawyer should be prepared to "scan" the output to identify relevant documents. This can usually be done quickly, since irrelevant documents can often be identified by reference to the "key words" at the beginning of the headnote. This scanning process normally reveals the peculiar problems created by the use of words in legal contexts.

Terminology

Though the common law has developed an endemic vocabulary, that vocabulary is neither static nor consistent. The problem which arises when words have common legal and non-legal meanings is obvious. However, the imprecise or variable meaning of many legal terms poses a larger problem. For example, the phrase "false imprisonment" appears

to embody a fairly precise idea. But a hastily formulated search of "false imprisonment" might produce disappointing results, assuming that the searcher seeks cases dealing with false imprisonment by persons who are not police officers, because cases involving police arrests would create a high degree of irrelevancy. To decrease irrelevancy, the same search might be re-formulated to exclude certain words likely to occur in headnotes describing police arrests (i.e. exclude the word "police"). This would probably reduce the number of irrelevant headnotes retrieved, but it would also increase the possibility of missing some relevant cases in which the excluded word appeared extraneously, or by way of distinction.

Further difficulties are created by inconsistent citation practices, especially with regard to statutes. For example, a reference to section 452 of the Criminal Code, dealing with the duty of police officers on an arrest without a warrant, could appear in a headnote in any one of several forms:

Example: s.452, s.452(1), ss.451, 452, 453, (etc).

Because the computer treats a string of numbers in the same way as it treats a word, one may simply search 452 or 452(1), to retrieve all documents containing either of these numbers. This approach would eliminate the problem of trying to search a phrase containing all the possible variations on "section". One might phrase the search to require that the term "criminal code" be present in the same headnote. The search would then be:

"criminal code" & 452 452(1).

This search could be further refined to include headnotes which abbreviate the word "criminal":

criminal cr. & code & 452 452(1).

These and similar refinements increase comprehensiveness and reduce noise. However, an unpredictable degree of noise would be inevitable, because the number 452 will occur in contexts other than statute citations (i.e. as the page number in a case citation at the end of a headnote).

Generally, the problems created by terminology are not insurmountable, and they will seldom render a search infeasible. They do, however, decrease the comprehensiveness and relevancy of some searches.

Coverage

A glance at the Appendix will reveal the following:

- (a) Not all Canadian statutes, regulations, and reported decisions (headnotes) are on the system;
- (b) Date bases contain only headnotes, not the full text of reported decisions;
- (c) Headnote data bases (with the exception of the Supreme Court Reports) are not complete;
- (d) Headnote data bases are updated at intervals which average about one month.⁶

These limitations can be corrected in time, with the entry of new data bases and the extension of existing ones. Full-text entry of reported decisions would further increase possible access points and eliminate distortions created by headnote editors, but at the same time, the additional volume of words would probably result in increased noise and irrelevancy. Experiments with full-text searching in the United States has produced conflicting reactions.

HOW THE COMPUTER CAN ASSIST IN PRACTICAL LEGAL RESEARCH

The preceding observations explained in outline how the computer works. Several observations may now be made regarding the practical application of computerized legal information retrieval.

A study commissioned in 1981 by the Canadian Law Information Council⁷ involved the research of ten actual legal problems submitted by practising lawyers. Each of the ten problems was first searched in the traditional manner and then a combined computerized-book research method was used. The study claims a 72.3% time saving for the combined computerized-book method. The traditional method found 88% of the relevant cases and the combined method found 74%. The study implies that the 14% deficiency in relevant cases retrieved by the combined method was due to the fact that headnote data bases are at present incomplete in the periods they cover. It appears highly likely that if more time had been spent supplementing the computer searches with book research, the deficiency could have been reduced or eliminated, without a drastic reduction of time saved.

The 1981 study reveals several important facts. First, computer-assisted legal research can substantially reduce the time spent on research:

⁶Exact update information can be obtained with each search.

⁷Iosipescu, Michael and Yogis, John: *A Comparison of Automated and Manual Legal Research: A Computer Study*, (Ottawa: Canadian Law Information Council, 1981).

total time spent on book research was 116 hours compared to 31.12 hours for the combined method. This time saving involved a sacrifice of fourteen percent of the relevant cases. Significantly, neither approach found all the relevant cases: some cases located on the computer were not found in manual research, and *vice versa*.

Second, both methods produced many irrelevant cases, necessitating a process of examination and selection. Approximately the same number of irrelevant cases was retrieved by both methods. The familiar criticism that the computer retrieves too much irrelevant material seems to be based on unreasonable expectations: a rough count of the number of "false leads" experienced in a typical manual search would likely prove flattering to the computer.

Third, the computerized method usually requires a manual preparatory search. Thesauri and dictionaries are often used to identify synonyms; textbooks or periodicals may be consulted to isolate key concepts which are then reduced to concrete words or phrases; the headnotes of readily available cases on point may be examined to extract key words or phrases which are likely to recur in the headnotes of similar cases; the names of leading cases may be noted and searched alone or in combination with fact-words or issue-words. The preparatory manual search can be as extensive as time or resources permit. However, in the usual case the preparatory search does not require access to an extensive collection of materials because sources consulted are representative only.

The computer supplements, but does not replace, manual research. There are no rigid rules governing the precise combination of the two methods; this will vary according to the issues involved and the individual lawyer's research habits. The unique contribution of computerized legal information retrieval would appear to lie in the fact that it offers the lawyer a new and different method of accessing the law. Its practical time-saving advantages can provide the busy practitioner with a viable method to improve general research habits, and this factor alone seems to promise continued intensive use and development in the future.

M. McGUIRE*

*B.A., LL.B. (University of Ottawa), M.L.S. (University of Western Ontario). CLIC Service Centre Coordinator, Reference Librarian and Instructor, University of New Brunswick.

APPENDIX A

DATA BASES

- NOTE: The time span which each data base covers is listed after the data base names. Data bases appear to be updated on a monthly or bi-monthly basis, with the exception of SCR which is updated less frequently. The date of the most recent update can usually be obtained at the time of the search.
- ACWS — All Canada Weekly Summaries. Concise summaries of all civil cases received by Canada Law Book Ltd. January 7, 1977 — last update.
- APR — Atlantic Provinces Reports. Headnotes from the New Brunswick Reports (2d series), Nova Scotia Reports (2d series), and the Newfoundland and Prince Edward Islands Reports. Published by Maritime Law Book Co. Ltd. and sponsored by the Canadian Law Information Council. 1969 — last update.
- CCC — Canadian Criminal Cases. Headnotes of the 2d series, edited by Canada Law Book Limited. 1971 — last update.
- DLR — Dominion Law Reports. Headnotes of 2d series and of 3d series, edited by Canada Law Book Ltd. September, 1955 — last update.
- FCR — Headnotes of the Federal Court of Canada Reports. 1971 to last update.
- RSC — Revised Statutes of Canada. Office consolidation of acts in force at December, 1981. Prepared by the Department of Justice.
- SAC — Statutes of Alberta Citor. References to Judicial decisions concerning Alberta Statutes. Edited by the University of Calgary Law Library and sponsored by CLIC.
- SBC — Statutes of British Columbia, prepared by the Ministry of Attorney General. Amendments affected by proclamation, and filed with Registrar of Regulations after October 13, 1981 not included in text of acts, but references to them are in historical notes at end of each affected section. Also, text of unproclaimed amendments is now searchable and contained in a document entitled "Unproclaimed Amendments". Note: 1980 Amendments to Consumer Protection Act, and also to Cattle Horn Act and Livestock Public Sale Act (see Livestock Brand Act) not included in the data base.
- SCR — Headnotes of the Supreme Court of Canada Reports: 1876 — last update.
- SMC — Statutes of Manitoba Citor. Reported and unreported judicial decisions since 1970. Sponsored by CLIC and the Law Society of Manitoba.

- SNB — An office consolidation of the New Brunswick Statutes complete to March 10, 1981. Prepared by the New Brunswick Office of the Attorney General.
- SO — Statutes of Ontario. An office consolidation to January 1, 1981. Incomplete and in preparation.
- SOR — Statutory Orders and Regulations. A selected data base of regulations prepared by the Department of Justice.
- TAR — Tax Advance Rulings. Income tax rulings published by the Deputy Minister of National Revenue for Taxation of Canada.
- WCB — Weekly Criminal Bulletin. Concise summaries of all criminal cases received by Canada Law Book Limited. October 27, 1976 to last update.
- WWR — Western Weekly Reports. Headnotes from 1968, Volume 62 to last update. Published by the Carswell Co. Ltd. and operated for the Canadian Law Information Council.