

INSTITUTE OF CURRENT WORLD AFFAIRS

RFS - 3 THE IDEAL LEGAL INFORMATION
RETRIEVAL SYSTEM

P. O. Box 14246
Nairobi, Kenya
19th April 1973

Mr. Richard Nolte
Executive Director
Institute of Current World Affairs
535 Fifth Avenue
New York, N. Y. 10017
U.S.A.

Dear Mr. Nolte:

In my last newsletter I discussed my M.Sc. course at Birkbeck College and my experience with adult education. As part of the requirements for the degree, I was asked to write a thesis. This thesis was on legal information retrieval systems.

The problems involved in designing legal information retrieval systems are similar to those in other fields with information retrieval needs. So for your consideration I have edited the last chapter of my thesis. I hope you find this extract interesting and informative.

Sincerely,



Roberta F. A. Spitzberg

1.

The field of law is generating more information than any practitioner can possibly contend with. Particularly in the United States, statute law is being generated at an amazing rate. The number of cases tried is also increasing geometrically. This huge volume of information leads to inefficient searching techniques and probably inadequately researched cases, thereby increasing the number of cases that must be retried or appealed. The most extreme instance is that this legal information explosion could lead to wrong (legally invalid) decisions.

The problem of this vast increase in legal materials is compounded by the fact that old legal cases are just as relevant to a lawyer as new. The basic legal principle of the common law countries is stare decisis (loosely translated the decision governs). Stare decisis is the principle of legal precedent: previous court cases are used as the basis for further decisions.

To illustrate how great is the increase in legal materials it is only necessary to consider that the Harvard Law Library has over one million books and needs over $4\frac{1}{2}$ miles of new shelf space to house its new acquisitions each year. It is obvious that this vast amount of material cannot be adequately used without help.

Before discussing the ideal legal information retrieval system, we must first consider whether retrieval of legal materials differs from retrieval of any other type of text. A first reaction is to answer that legal information retrieval is merely a subset of document retrieval. This is true, but the law does pose certain problems that do not exist in other areas.

Any user of an information retrieval system wants to receive from a search a list of documents selected by the retrieval process (citations) that contains only the documents relevant to this research problem. In general, the omission of one relevant document would not cause any disastrous consequences. This is not necessarily true for the law. The omission of a case can literally be a matter of life and death. Therefore, the retrieval of all relevant documents is essential for law.

Another important consideration is the availability of the system. Every lawyer should have access to a system, but should non-lawyers have access as well? At present, at least in the United States, any individual can use legal materials in public libraries. But many lawyers do not want non-lawyers to have access to a legal information retrieval system. Other lawyers view the computer itself as possibly being involved in unauthorized practice of law. The problems of unauthorized practice can only be solved by the lawyers themselves, but it is important to realize that a problem exists.

The ideal legal information retrieval system must have a comprehensive data base. In order to determine what should actually be included, we must first consider whether we are discussing local, national, or international systems. In order to simplify the following discussion we will consider national and international systems as two separate ideals.

Is a national system a worthwhile ideal? The answer to this question must surely be yes. The data base for a national system in the United States, for example, would have to include all state and Federal law. This may not seem feasible, but such a data base is actually planned by Mead Data Control for its LEXIS System. A national system could not practically contain all local ordinances. These could be incorporated into a system maintained by local bar associations, if the bar associations deem local ordinances a necessary component of a computerized legal information retrieval system. A national system for common law countries other than the United States would include all statute law currently in force, case law, administrative and regulatory law.

The data base should consist of the full text of all documents, supplemented by index terms. All new material should be indexed by the judge, attorney, or legislator who authored the document. Existing material should be indexed by experts, perhaps members of the bar association in the jurisdiction in which the material originated. The index terms for cases should include bibliographic information such as author of document, jurisdiction, court, factual information and points of law. Other information that would appear as index terms might be West Key Numbers (a system of indexing developed by the West Publishing Company) and important cases cited in the opinion. Statute law would be indexed by concepts of law and related statutory sections as well as bibliographic terms. Indexing should be viewed as an augmenting of material. The inclusion of West Key Numbers would be important during the transition from manual to computer research as most lawyers use West in their research now.

A decision must be reached by legal publishers to prepare all material for publication using photocomposition techniques or any other method that will produce the text in machine-readable form. Any reissuance of material should be published in this manner. A program must be started by bar associations and organizations such as the U.S. Government Printing Office and Her Majesty's Stationery Office to convert existing legal materials to machine-readable format.

The query language of a legal information retrieval system should be specifically designed for the law. Facilities should be provided so that a lawyer can find all statutes where a term is defined; such a facility is available in the STATUS system of the UKAEA. Searches might specify index terms only, non-common text words and phrases only,

or a combination of index terms and non-common words. An easy means of searching on phrases should be provided, rather than relying solely on metric operators for phrase representation. But both Boolean and metric operators should be part of the query language.

The searcher should be able to obtain the count of documents found relevant at any point in the search process. This count information would serve to aid the user in deciding if his search request needs to be refined or expanded. The user must be able to modify any part of his search request. The user must also be able to restrict the search to any subset of the data base. Lists showing the number of times non-common words appear in the data should be available to help the searcher choose search terms.

The above search capabilities could only be provided in an interactive system. A legal information retrieval system must be interactive. Only the researcher can adequately define his own problem. Only the researcher can decide if, and how, his original request needs to be modified. Intermediaries should be avoided and the only effective way of doing so is to have complete interaction between the user and the system.

The system should have a thesaurus. The approach used by the U.S. Department of Justice JURIS system provides the most flexibility. The thesaurus, whether manually or machine generated, should be stored. The user can display the synonym classes of any search terms and determine which synonyms and grammatical equivalents he should use to expand the search. Automatic generation of grammatical variants should be optionally provided. It is important that search term expansion by inclusion of synonyms and grammatical variants be user controlled. The default option should be no search term expansion. Although this may make search framing more difficult for the new user, search term expansion must be controlled by the user who alone is able to analyze the problem and determine the best method of articulating it.

The ability to obtain KWIC (Key-Word-In-Context) output at any point in the search is also strongly recommended. A short KWIC would serve to aid the user in determining if his search is framed correctly. This compact KWIC would enable the user to see the search terms he has chosen in the context of the retrieved documents.

Other forms of output would be citation lists and full text. A count of documents currently deemed relevant should be available at any time. This would serve to give the user some indication of the quality of his search term selection. The full text of any document should be available as an output option, but it should not be printed on-line. The most efficient way to handle full text output would be by means of a computer

controlled microfilm reader-printer. This would enable the user to scan the full text of retrieved documents and obtain hard copy if necessary.

Citation identification and citation networking should be a part of any legal information retrieval system. Some systems do use KWIC indices to "simulate" citation networks, but no actual identification or networking facility is provided. Citation identification has reached a level of sophistication to justify its inclusion. Networking has been used by NASA, to name just one example, for bibliographic citations. To most lawyers, citation networking is an integral part of legal research. Any legal information retrieval system without this capability is incomplete.

Since we are discussing a national system, portability may not seem to be an essential feature. Yet we would want the system easily adaptable for use in other countries. Although Niblett and Price have proven through STATUS that an information retrieval system can be workable when written in ALGOL and FORTRAN, we know that the use of these languages leads to inefficiencies. PL/I would be a better choice of higher level language, but we would then be restricting ourselves almost exclusively to IBM equipment. Portability may seem to be desirable, but it must not be given too high a priority. The only specific hardware recommendation I am prepared to make concerns the type of terminal. The terminal should consist of a CRT with a character printer and a computer driven microfilm reader-printer. The use of microfilm causes no more updating problems than the widely used legal loose leaf services.

Any national system needs the support of the national bar association as well as local bar associations. It is the responsibility of the American Bar Association and the Law Society to investigate legal information retrieval systems and authorize one system to be used nationally. This is what both the Ohio and New York Bar Associations have done. Since MDC is already working on a national system, it would seem that the American Bar Association would choose it, and this may not be a poor choice. The ABA should, however, make intelligent recommendations and proposals to MDC as the Ohio Bar Association did when OBAR was being developed. The involvement of the bar is essential. The bar should be the administrator of a legal information retrieval system. This has proven itself to be workable and desirable by the experience of the Ohio Bar and MDC.

The best choice of an administrator for an international system is the World Peace Through Law Center in Geneva, Switzerland. This international organization has already recognized the need for an international system. The World Peace Through Law Center has proposed that an international base consist of 1) treaties and international agreements, 2) international court decisions, 3) national high (supreme) court decisions, 4) international organization conventions, declarations, regulations, and

decisions, 5) statutes, 6) ordinances and charters of cities, 7) model codes, and 8) customary international law.

An international data base would necessarily be written in many languages. Search requests would also be written in many languages. The experimental work of Salton at Cornell and Mackaay and Fabien at the University of Montreal have shown the feasibility of using full text retrieval on a multilingual data base. Salton has had some success using a multilingual thesaurus. It is therefore practical to propose that the ideal international legal information retrieval system will have all the features in the ideal national system plus some additional features. The data base would be the one described above. All documents would be the full text in the original language. A multilingual thesaurus would be available to enable a searcher to obtain relevant documents written in any language of the data base. The searcher should be able to restrict the search to documents of any country or to any type of document. There is no doubt that such a system would be invaluable.

At present, no existing system possesses all the features of the ideal. Yet all the features discussed are practical and capable of being implemented today. The one thing that is lacking in order to make the ideal an operational system is a coordinated effort by lawyers themselves. Without their cooperation and initiative, the ideal system cannot become an actuality.

Received in New York on May 2, 1973