Book Chapters   Faculty Scholarship

1971

Computer Systems for Research

Layman E. Allen
University of Michigan Law School, laymanal@umich.edu

9

Computer Systems for Research

by Layman E. Allen*

Attention in this panel is shifted from a consideration of the legal implications of modern communications technology to an examination of its potentials for improving the communication network in those affairs of men that we call the law. Our discussion will be somewhat of a smorgasbord—a selection from among many potential applications within the legal communication network. My introductory remarks will touch lightly on research in law. Others on the program will broaden the discussion to include the courts, the practicing lawyer, and, in part, the general public.

Characteristics of Legal Communication Network

The legal communication network today is characterized by two features. Any communication network in this century is marked by a division between the extent to which there is a man involved and the extent to which there is a machine involved. And, in terms of emphasis at this stage of things, at least within law, the emphasis is heavily upon the man communicating messages and relatively less upon the machine. The interesting question is, What is going on within this network that is amenable to being handled by machine and what, among those things, is it wise to do that way?

---

*Professor of Law, The University of Michigan Law School, Ann Arbor, Michigan.

A second and limiting characteristic of the legal communication network arises from the fact that we restrict ourselves to the English language as virtually our sole means of communication, which limits the use we can make of machines. It may be that we will begin to use some slightly modified or even radically modified languages in our communications.

Uses of Machine Retrieval

It is useful in legal research, for purposes of isolating some of those things that may be done wisely by machines, to draw a distinction between document retrieval and information retrieval. By "document retrieval" I mean the process by which you identify and locate a document. "Information retrieval" refers to something beyond that, to those events that occur after you obtain a document. Information retrieval begins when a retrieved document is read and related to a problem with the goal of finding a solution to that problem. This latter process may be amenable, in part, to being done by machine. To date, however, the work that has been done has focused almost entirely upon the former process—the process of document retrieval.

It is not surprising that efforts have tended to parallel the traditional tools which we have been using for years for the purpose of locating a document. Although the methods are familiar, the particular names which information technologists use are not as well known. Briefly, these methods are called citator systems, descriptor systems, and key word systems.

In a citator system, such as Shepard's, the index of relevance of a document is the fact that it has been cited in another document. It is presumed that, for that reason, it may be useful or relevant for another purpose.

A familiar descriptor system is the West system, containing some eighty thousand categories with which a human indexer has associated a particular document. He has made a judgment that this particular document is related to this or that category. When we want to find that document, we think of the word that designates that category and look it up. With so many categories, it is not surpris-

ing that we frequently run into difficulty trying to guess how a particular indexer would classify a particular document. The index of relevance in descriptor systems is the association with a particular category, a particular word that may or may not be mentioned in the actual target document that is being sought.

A key word system uses yet a third index of relevance: the occurrence of a word within the document. The key word systems range all the way from what is called "full text searching," which refers to searching by the use of all except two or three hundred most common words, to a system very similar to the descriptor systems, using only a small fraction of the words in a document to retrieve that document.

On the information retrieval side of legal research, there is now well over a decade of highly relevant experience. However, despite significant research efforts at automatic translation, the quality of the automatic translation which we are able to do today is not of the highest quality. We certainly can discern what the documents are about, but we have encountered great diffculty in getting a machine to show the precise content of a document. It has become quite clear that our natural language is a much more subtle and formidable tool to deal with automatically than had been fully appreciated when these efforts began. Particularly, it is the syntax of natural language that is rather difficult to handle automatically. If we were able to simplify the language, in the sense of imposing some standardization or normalization, we could begin to do some of the analysis, some of the information retrieval, automatically by machine. We do know how to deduce automatically by machine, if we could find a way to change the language expressing our ideas to a form that could be more adequately processed.

One such form, particularly useful for the expression of legal norms, is the pattern of logical entailment. This pattern involves expressing norms in the form: if a specified set of conditions is fulfilled with various relations holding between them, then certain legal consequences follow. Such ideas can be expressed in a way that rather closely follows ordinary English prose.

Some work in this vein has been going on at the University of Michigan with respect to the Internal Revenue Code. The ideas embodied in Sections 354 and 357 of the code (those dealing with corporate reorganization) have been re-expressed in a form which allows some deduction to be done automatically by machine. It was the impression of several of us that this new form was easier for us as human beings to work with. We administered a test to a group of law students here. We gave the students a version of Section 354 that was cast in this normalized form, a second version exactly as it is now, and then a series of questions which asked about the effects of the statute. Using measurements of speed and accuracy as tests, we had instructive results. We anticipated that the group would probably be able to work faster, and they did, on the average of about twenty percent faster, in dealing with the normalized form. The most surprising result was that the order of difference in accuracy was better than forty percent in working such simplified problems. Because the problems were simplified, there were answers that could be evaluated in terms of correctness and incorrectness.

Thus, one of the side benefits of work, in trying to see what more can be done by machine, is improvement in our manual processes for retrieving and processing information. What may be an output of this effort, which was originally directed toward more automatic processes, is better looseleaf files that begin to have more of the characteristics of a dictionary. In a dictionary we have a very effective "homing-in" procedure. That is, through the alphabetical order of letters used in words, wherever we land in a dictionary we know how to "home in" on where an item will be in the file. In organizing a looseleaf file according to the logical properties of the sentences used to express the ideas, we can similarly "home in" on a location in the file where an item should be, if it is there at all.

Within the kinds of things that lawyers do, there sometimes are genuinely scientific problems that ought to be tackled by a scientific method. The nature of legal language happens to be one of

them, it seems to me. If I am correct, then we are faced with the simple empirical problem of evaluating competing ways of expressing ideas, in terms of our own ability to deal with them. As John Horty remarked, "Once we open the door to our profession and let in the mathematicians, logicians, computer scientists and others to take a look at the legal profession, the whole process of the administration of justice is never going to be the same again."[1] And the words are having a truer and truer ring with the passage of each year.

1. HORTY, *The "Key Words in Combination" Approach,* 1962 MODERN USES OF LOGIC IN LAW 54, 64.