Education in Information Science

This report discusses the various approaches being taken to education in information science. It develops specific definitions of "information," "information system," and "information science" which serve to rationalize these approaches. The result is a common core of interest to which the variety of curricula can be related. The differences in curricula are exhibited in terms of the relative priority they will assign to different course offerings.

R. M. HAYES

Institute of Library Research
University of California
Los Angeles, California

• Introduction

Education in information science has been defined in such a variety of ways that it is almost impossible to extract a common thread from them. Just to list a few:

1. Some schools, all of them library schools, have identified information science with the use of computers in libraries. Thus, such schools will imply that they have an "information science curriculum" if they have added one or two courses on "data processing in the library." The confusion created by this definition has been compounded by federal legislation in its references to "library and information science," which has usually been interpreted in this sense.
2. Some schools, all of them library schools, have identified information science with "science information" and usually regard it as synonymous with "documentation." Such schools will add courses on "indexing and abstracting," on "the management of information centers," and on "mechanized information retrieval" as their curriculum on information science.
3. Some schools, usually engineering schools, have identified information science with computer science, usually with emphasis on the use of computers for processing natural language, for "question-answering," perhaps for heuristic programs and "artificial intelligence."
4. Some schools, usually engineering schools, have identified information science with communication theory and regard it purely as a sub-set of their existing curricula.
5. Some schools, of a variety of kinds (medicine, urban planning, business administration, engineering, librarianship) have identified information science with the design of information systems in their own specialized fields of interest (with patient-monitoring systems, with urban data banks, with management information systems, with command and control systems, and with library systems). Usually, but not always, they will call their curricula "information systems" curricula and include courses on system design and on the application of computers to their specialized field.
6. Some schools regard information science as a discipline in its own right, which, while applicable in many fields, has its own problems of research interest, its own methods of research, its own "discipline." Such schools will include both theoretical courses (drawn usually from the formal disciplines of mathematics, logic, and perhaps linguistics) and applied courses (from fields like psychology, engineering, or microbiology).

Why such a diversity? Is there any common thread which will tie all of these together and provide the means for rationalizing one curriculum with another? The purpose of this report is to answer these two questions—to identify the reasons for diversity and to define a common thread. To do so, however, we must define three things: (1) information, (2) information systems, and (3) information science. These definitions admittedly will be highly personal and will not fit the ways in which these terms are generally used. But everything—including the curricula listed above—indicates that no definition will satisfy everyone.

• Definitions

Information

"Information" has had a variety of meanings. At the one extreme, engineers will identify it with transmission over communication lines and will measure it by the statistical properties of signals. At the other extreme, philosophers may well claim that it is a question with which they have struggled for years; don't presume to have one answer. Some have identified information with recorded facts; others, with the content of text; others with the experience stored in the human mind.

The following is a precise and operational definition of "information":

> *Information* is the data produced as a result of a process upon data. That process may simply be one of *transmission* (in which case, the definition and measure used in communication theory are applicable); it may be one of *selection;* it may be one of *organization;* it may be one of *analysis*.

Since information is itself recorded as data, it can be subjected to subsequent processing. The issue is, therefore, really, one of the degree to which data constitute information, as a function of the complexity of the process to which it has been subjected. Hence, measures of information must recognize not only "amount" but "complexity."

An important point about this definition is that, given it, information like life cannot exist independent of the processes which produce it. Although some of its properties may be subject to investigation independent of the means of performing those processes, the important and interesting ones cannot be. That means that "information" can be studied only in the context of specific "information systems."

INFORMATION SYSTEM

Consider any complex phenomenon (i.e., "system") of interest. If one wishes to study it, one may be concerned with a variety of things about it—its physical structure, its cybernetic response to environment, its metabolic balance, its information processing. Thus, a person can be viewed as a complex of bones and muscles, capable of performing mechanical tasks; as a chemical factory processing ingested food, water, and air and converting it to energy; or as a thinking human being, taking in sensory data, making decisions, and controlling its physical and chemical structure. A library can be viewed as a physical structure, with physical records and a mechanical flow of materials; as an administrative organization, with the assignment of people to a variety of tasks; or as an information processing institution, taking in data and providing it out again in response to requests.

We study such systems by using methods of research appropriate to them—biology for living organisms, organization theory for administrative systems, etc. The results of study are a variety of models, or scientific theories, which we then use to explain their behavior and predict future behavior.

But among the aspects of interest in a system are those which we may want to identify as information processing. If those are the ones on which we focus, the system then becomes an information system. Hence,

> An *information system* is that set of aspects of a general system (a natural phenomenon, a physical construct, or a logical construct) which are identified as information-producing.

• Information Science

> That leads to a very natural definition of information science: *Information science* is the study of information-producing processes in any information system in which they may occur.

However, although this indeed is the definition which we will use, it implies something which may not be true, viz., that information science exists as a separate discipline.

In fact, any real system can be studied only by use of the scientific methodologies appropriate to it. And that includes the study of the information-producing processes embodied in it. For example, RNA and DNA can be studied for the information processes which they embody and, as such, are information systems of vital interest to information science. But it would be impossible to study, in any real sense, the means by which they transmit, select, and even analyze "data" (represented by the various configurations of amino acids) without the use of microbiology.

This means that while "information science" may in principle be concerned with pure analysis of processes, in reality it cannot be separated from the methodology of specific disciplines. It is therefore more appropriate to talk in terms of "information science in genetics," or "information science in social theory," or "information science in documentation" than to talk of "information science" in isolation from specific systems.

• Contexts

What then are the "systems" with which information science is concerned? What are the methodologies for scientific study with which it must work? Table 1 summarizes some relevant examples, a few of which will be discussed in detail.

THE COMPUTER

The computer has been an especially important context for information science, for very clear reasons. The *raison d'être* of the computer is its data processing capabilities. Furthermore, those capabilities are very well defined and can be measured in very precise ways. The computer is therefore a relatively predictable system to study.

There now exist a large number of departments of computer science which bring to bear the combined methodologies of engineering for study of the characteristics of computer hardware and of mathematics and logic for study of the characteristics of computer software. In general, if any one focus can be defined for computer

TABLE 1. Some systems with which information science is concerned

Systems	Science	Specific examples	Related sciences	Data form	Representative processes
Formal	Mathematics	Mathematics	Mathematics	Symbolic	
		Logic	Logic	Symbolic	
		Language	Linguistics	Sounds, printed	
		Computer software	Computer science	Symbolic	Branching
Engineering	Engineering	Computer hardware	Computer science	Pulses, states	Switching
		Communication	Information Theory	Pulses, states, codes	
Computer-based information	Systems Methodology	Management information systems	OR., Management science	Numerical	PERT
		Urban data banks	Urban Planning	Numerical	
		Command and control	Military science	Numerical	
		Configuration Management	Engineering	Numerical	
		Mechanized Lit. Retr.	Documentation	Textual	
Recorded data	Systems Methodology	Libraries	Library science	Print, microform	Cataloging, reference
		Information Centers	Documentation	Print, microform	Indexing, analysis
		Science information network	Documentation	Print, microform	Meetings
Education	Psychology	Education	Psychology		
Social	Social Science	Economy	Economics		
		Politics	Political Science		
		Organization	Organization Theory		
Biological	Biology	Human brain	Neurology	Neurogenic	
		Human response	Psychology, psychiatry	Responses	
		Genetics	Microbiology	Amino acids	

science, it is that of "realizability." Thus, it should be expected to answer questions of "formal realizability" representing recursive function theory, "program realizability" representing the ability to produce an operating program for a specified task, and "pragmatic realizability" representing the ability of hardware to execute the program within specified time limits.

Because computer science has become formalized at the same time as information science, and because the computer is the most clear-cut example of an information system, it has been natural to identify information science with computer science. However, to do so limits information science unnecessarily.

COMPUTER-BASED INFORMATION SYSTEMS

As the computer has been used in an increasing variety of applications, the resulting computer-based information systems have become a fruitful spawning ground for information-science activities. They embody information-producing processes considerably more complex than those of the computer itself. The result has been the growth of "systems work" or "application study"—the body of techniques by which an organization is studied and alternative systems for information processing are designed and evaluated. However, the use of those techniques requires an informed substantive knowledge of the area of application, whether it be a business company, an urban community, a miltary command and control system, a scientific information center, or a library.

Since the primary focus of systems work is on the information processes in these areas of application of computers, it has been natural to identify information science with systems work on computer applications. Again, to do so unnecessarily limits the scope of information science.

Libraries and Information Centers

Libraries and information centers, as institutions, exist solely for the information functions they serve. Furthermore, the information processes they use—cataloging and indexing, for example—are relatively well formalized, although there are still large areas of their work which are not well understood. As a result, they provide ideal subjects for study by information science, and there has been a corresponding interest in libraries and information centers to utilize the insights which information science could provide. The extent of interest, on both sides, has been so great that some have identified information science with library science. But it seems clear that to do so unnecessarily restricts the scope of information science.

Social Systems and Biological Systems

That a social system or biological system can be regarded as an "information system" may seem somewhat unnatural, at least in the sense we will use here. But it is clear that each performs processes upon what we can regard as data (symbolic representations). For example, an elected official can be treated as a symbol of his constituency. There are clearly defined mechanisms for selection of such symbols and these constitute "information-generating processes." Similarly, the "genetic code" is simply arrays of amino acids, but these can be regarded as data (symbolic representations). Economics is concerned with processes upon symbols of capital; psychology, with processes upon symbols of response, etc.

Each of these is therefore a proper domain for information science to study.

• Curricula

With such an array of contexts within which information science would have meaning, it is natural to expect a comparable array of curricula. Here we will highlight four—those which at the moment have been most clearly defined and identified with information science. Table 2 presents a comparison of their different course emphases.

Computer-Science-Oriented Curricula

Since these tend to follow the pattern of computer-science curricula in general, all that need be said is to refer to the report of the ACM Curriculum Committee.

Table 2. Comparison of curricula. PR denotes a prerequisite course, 1 a required course, 2 a recommended course, and 3 an elective.

Representative courses	Orientation of curriculum: Theory	Computer	Systems	Library
Formal disciplines				
Calculus	PR	PR	PR	
Programming	PR	PR	1	2
Symbolic logic	1	1	3	
Recursive functions	1	1	3	
Linguistics	1	2	2	
Applied disciplines				
Statistics	3	3	1	
Operations research	3	3	1	3
Psychology	2	3		
Information theory	1	1	3	
Systems analysis		3	1	2
Methods of social research			2	
Computer-oriented courses				
Computer hardware	3	1	2	3
Compiler construction	3	1		
Data base management	3	1	1	2
Information retrieval system	1	1	1	1
Management information system		3	2	
Management-oriented courses				
Managerial accounting			1	3
Organization theory	3		2	
Information center mgt.			1	1
Service-oriented courses				
Sources of information			2	1
Catal., class., index., abst.	2	2	2	1
Documentation			2	1

Library-Oriented Curricula

These are the curricula which library schools concerned with "functional education" and preparation for practice have adopted. They tend to treat information science as simply the use of computers in support of present day operations in libraries and information centers, but including mechanized information retrieval.

Systems-Oriented Curricula

These are the curricula which emphasize the methodology of systems analysis as it applies in one or more institutional contexts. For example, such a curriculum in a library school would emphasize an integrative view of the library—management, systems and procedures, etc.—rather than the functional view of the practice oriented curricula.

Theory-Oriented Curricula

These are the curricula based on the premise that information science is a distinct discipline. They tend to emphasize the formal systems (mathematics, logic, language, and computer software), since the theory is most clearly exhibited there.