1980

DOCUMENT PROCESSING IN THE AUTOMATED OFFICE: IMPLICATIONS FOR MIS RESEARCH AND EDUCATION

Mary J. Culnan
University of Virginia

Follow this and additional works at: http://aisel.aisnet.org/icis1980

DOCUMENT PROCESSING IN THE AUTOMATED OFFICE: IMPLICATIONS FOR MIS RESEARCH AND EDUCATION

MARY J. CULNAN

McIntire School of Commerce
University of Virginia

ABSTRACT

Document processing is an integral part of the automated office. Because document-based information systems are composed largely of unstructured text as opposed to structured data, the techniques which have been successfully applied to the design of current database systems are not adequate for the design of office information systems. The paper identifies issues related to the storage and retrieval of office documents, and the management of information in the automated office. Strategies for broadening MIS research and education to address these issues are suggested.

INTRODUCTION

Document and text processing are integral parts of the automated office. The purpose of this paper is to identify research issues related to document processing in the office, and to suggest ways to broaden MIS curricula to address these issues.

Office automation has only recently become an important MIS issue. The initial conceptualization of such systems, however, precedes the widespread use of computers for business data processing. For example, the following quotation by Vannever Bush appeared in the Atlantic Monthly in 1945:

> Consider a future device for individual use which is sort of a mechanized private file and library. It needs a name, and, to coin one at random, "memex" will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.(3)

In 1958, Luhn described the design for a business intelligence system which "will utilize data-processing machines for auto-abstracting and auto-encoding of documents and for creating interest profiles for each of the 'action points' in an organization. Both incoming and internally generated documents are automatically abstracted, characterized by a word pattern, and automatically sent to appropriate action points."(11)

More recently, office automation has been defined by the EDP Analyzer as a "new structured way of handling business documents and person-to-person communication." The scope of the automated office includes formal and informal correspondence and communication, storage and retrieval of office documents, and links to various corporate files and outside information services.(14) At the organizational level, the automated office provides a way to support the corporate memory, which Morgan and Root described as the totality of information sources such as those mentioned above, for all the individuals in the organization. The corporate memory is viewed as an active system which both responds to user requests as well as initiating interactions with organizational members.(16) At the individual level, the automated office may be viewed as a terminal-based system designed to provide decision-makers with access to both intra-organizational and extra-organizational information.

While documents are central to the automated office, document-based information systems have largely been neglected in the MIS literature.(20) Rather, the majority of current management information systems consist of structured databases containing largely internal, often numerical data. The same techniques which have been successfully applied to the design of these existing information systems are not adequate for the design of effective information systems to support the automated office, as documents consist largely of unstructured text. The primary issue addressed by this paper, then, is the storage and retrieval of an organization's office documents. A

secondary issue addressed by this paper concerns the management of management information as the automated office has the capability to provide any individual manager with instant access not only to internal documents, but to a bewildering array of external information sources as well. While the importance of external information to organizational survival has been recognized in the literature, there is little evidence that the MIS field has developed techniques for managing these information flows.(8, 21)
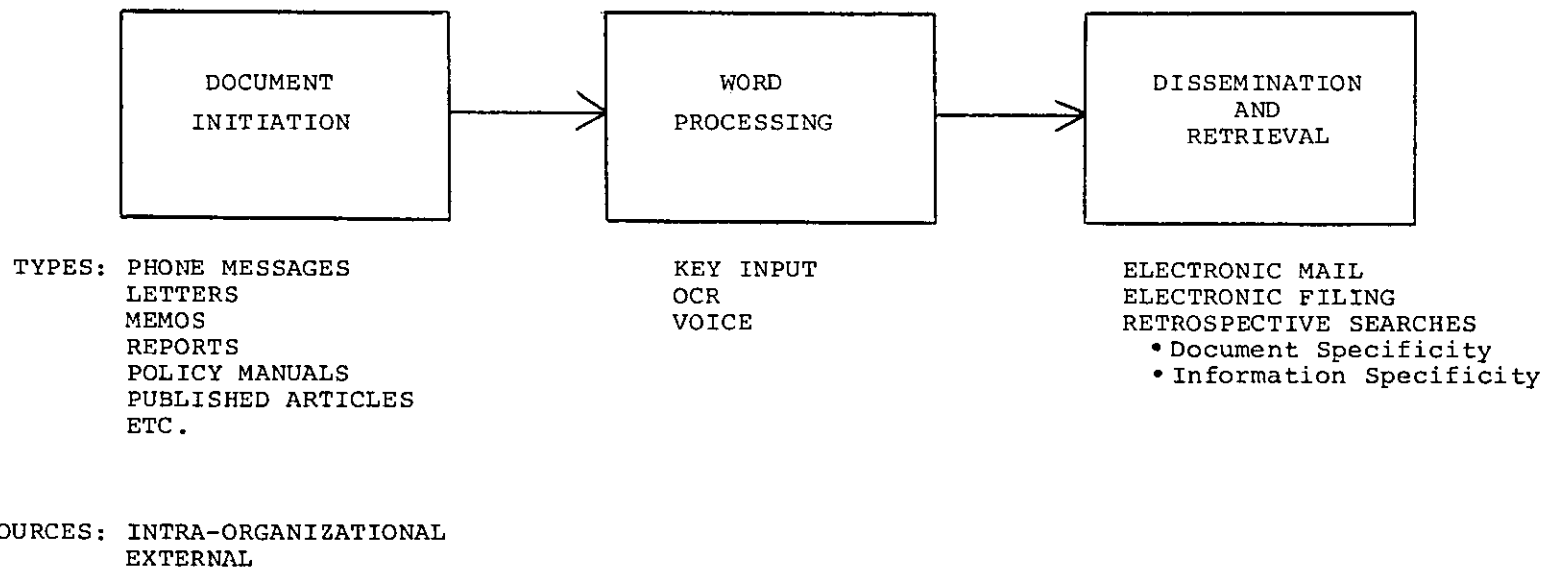
DOCUMENT STORAGE & RETRIEVAL

Document handling in the automated office consists of three phases: document initiation, word processing, and dissemination and retrieval. This process is represented graphically in Figure 1. Documents are created by an originator who may be an employee of the organization, or external to the organization. An office document may assume a variety of forms from a telephone message to a formal report. For the majority of intra-organizational documents, initiation and word processing may be combined into a single step. Once a document exists in digital format, it is disseminated to the recipient(s) by way of electronic mail, and/or filed electronically in the corporate memory for subsequent use. Further, individuals have the ability to tag a document for inclusion in their personal electronic files. Over time, individuals in the organization will retrieve documents from the corporate memory based on the information requirements associated with their jobs. In the case of retrospective searching, an individual may be seeking a specific known document(s) or seeking specific information from an unspecified set of documents.

On the surface, the process for handling documents would not seem to differ significantly from the process for handling data in most organizations. For example, consider the creation of a payroll rcord for a new employee. A record for the employee is initiated by the personnel department, based on the employee's application and subsequent data related to the person's employment. This data is subsequently converted to digital format by key entry, processed by a computer program, and filed in the company's data base. The data base may be queried retrospectively for a single record for a specific employee, or for an unknown group of records which fulfill some well-defined criteria such as skills codes, department, sex, salary, etc.

Despite the similarities between data and document processing, there are some significant differences which have important implications for the design and implementation of information systems to support the automated office. In the case of the employee payroll data, each record is a fixed size consisting of fixed-size fields; within every record, each field appears in the same location relative to the beginning of the record. While it is necessary to determine which fields in the record are to be used for retrieval at the time the data base is designed, once the system has been implemented, indexing occurs automatically by the data base software, in terms both of the key used to uniquely identify a given record such as employee number, and the secondary keys such as department, salary, date of employment, etc.

Document processing, however, presents a different set of challenges. First, documents come in a variety of lengths and formats--there is no such thing as a standard record. The information system must be able to process a heterogeneous set of documents ranging from digitized phone messages to formal reports. For example, prior to the design of the AMANDA system for intra-organizational document retrieval, filing cabinets, vaults, personal files and in-baskets were surveyed in two departments. Twenty-two categories of documents were identified including memos, reports, reprints and lab notebooks. (19) Second, documents consist of variable length words rather than fixed length, fixed location fields. While certain "fields" may be common to most documents, such as creation date or document originator, it is unlikely that each of these fields will be located in the same physical location relative to the beginning of the document. The key used to uniquely identify a given document may not be part of the document, but may instead be a number assigned by the system or the word processing operator. Third, the terms used to identify a record in the personnel data base are largely standardized, many may be numerical, such as a department number or a salary rate. For the text portions of these records, there may be only one spelling for each field such as an individual's name. In a document, different individuals may use different terms to represent the same concept. This will present little problem for personal files where a single individual may be responsible for tagging documents both for storage and retrieval. It presents a great potential problem in the case of the corporate memory where a document may be indexed and stored by one individual, and retrieved later by others, who may have had no contact with the

```
+------------------+          +------------------+          +------------------+
|                  |          |                  |          |                  |
|    DOCUMENT      |   ===>   |      WORD        |   ===>   |  DISSEMINATION   |
|                  |          |                  |          |       AND        |
|    INITIATION    |          |   PROCESSING     |          |    RETRIEVAL     |
|                  |          |                  |          |                  |
+------------------+          +------------------+          +------------------+
```

TYPES: PHONE MESSAGES   KEY INPUT   ELECTRONIC MAIL
        LETTERS   OCR   ELECTRONIC FILING
        MEMOS   VOICE   RETROSPECTIVE SEARCHES
        REPORTS   • Document Specificity
        POLICY MANUALS   • Information Specificity
        PUBLISHED ARTICLES
        ETC.

SOURCES: INTRA-ORGANIZATIONAL
          EXTERNAL

167

FIGURE 1

MODEL OF DOCUMENT PROCESSING

document(s) at the time of creation and original dissemination.

It is doubtful that any of the existing office automation systems, prototype or commercial, have successfully met the challenges presented by document processing, yet document retrieval serves as one of the most important capabilities provided by the automated office.(23) A recent study by IBM found that managers ranked retrieval as the third most frequently used prototype function, after mail queue and schedule calendar. Retrieval was ranked fourth by non-managerial professionals.(7) Many of the current systems utilize a "file folder" approach which provides only limited access to documents and in some cases, may represent mere automation of an organization's manual filing system. This approach may be described as "have solution, will travel," that is, these systems were developed based on techniques already used in other systems without considering either the possibilities afforded by the new technology or the decisions which the information system is to support.(15)

While the MIS field's experience with document processing is relatively limited, the field of library information science has more than a decade of experience with very large, document-based information systems. The majority of these systems are both bibliographic and reference-based meaning they serve as an index to published literature rather than actually containing the full text of the entire document as is likely to be the case in most office systems. It is likely, however, that the MIS field can benefit substantially both from the research and practical experience associated with these systems.

The majority of these bibliographic systems exist as commercial utilities. The database supplier acquires a body of literature pertaining to a given subject, e.g., business, chemistry, biology, engineering, agriculture, etc. A record is created for each document. Searchable fields are defined for each document, such as author, title, accession number, corporate source, date, document type and journal. In addition, the majority of these systems include subject terms which are assigned by an indexer with subject competence. The subject terms are normally selected from a thesaurus in order to ensure a high degree of standardization. The database may then be searched by subject terms, and/or by one of the searchable fields; in addition, all significant words in the title and abstract are indexed allowing for full text searching using an unrestricted vocabulary.

Much of the research related to the evaluation of retrieval strategies for these systems was conducted on databases in the fields of science and technology. These studies found first, that retrieval strategies based on assigned subject terms are likely to be less effective than those based on a combination of searchable fields including natural language terms from titles and abstracts.(4) Recent research indicates, however, that these findings may not be directly transferable to the retrieval of business documents. For example, a recent study conducted for the Office of Management and Budget performed a syntactical analysis of a sample of fifteen contractor cost performance reports from three government agencies. The study revealed that 427 different headings were used to identify 294 elements of numerical data; 343 of these headings were unique to a single agency. As an aside, it is interesting to note that most of the analysis was performed using a home computer.(13) While the capability exists for full-text retrieval of office documents, this study suggests that the use of an uncontrolled vocabulary may mean that potentially relevant documents will not be retrieved if the document's originator uses a different vocabulary than the searcher to express the same concept.

Given this background, then, what are some of the major issues related to the storage and retrieval of office documents that need to be addressed? Specific research strategies for dealing with these issues, and the implications for MIS education will be discussed in the final section of this paper.

1. What is the structure and the format of the office document? It is likely that document formats will vary with different types of documents. What searchable fields should be used in order to facilitate retrieval? For example, suppose a manager wished to retrieve all correspondence from a given customer. If the search were performed strictly on the content of the document, the results might also include correspondence to the customer as well as correspondence about the customer.

2. What is the vocabulary that is used in office documents in a given organization? What synonyms are used on a regular basis? If the vocabulary of an organization is fairly standard, it may be feasible to construct a thesaurus to be used in document indexing.

3. How shall documents be indexed, particularly for storage in the corporate memory? Should full-text be

used, or a controlled vocabulary, or some combination of the two.

4. Who should be responsible for the indexing of office documents? How much responsibility for document indexing may be assumed by the computer? In a number of current systems, the assumption is that indexing will be performed at the point of origination, by either a professional or a clerical person. (For example, in the AMANDA system, descriptors are assigned by the document originator using a thesaurus which was developed by the system's users.(19) Document recipients may also have the option of indexing a document for their own files. In order to enforce a minimum level of consistency, it may be that organizations will establish a new position of indexer. This person will be responsible for assigning descriptors to all office documents using a controlled vacabulary. In addition, individuals users will have the option of indexing documents for their personal files.

   Further, office systems are designed to support the goals of a specific organization. As a result, it may be necessary to index documents for facets which are not part of the document's content, but which indicate instead that the document is analogous to some context. For example, an article may suggest a technology that could be transferred to an entirely different problem being considered by the organization, or a letter may have legal implications which are not implicit in the text, but should be noted for future reference. In these instances, human judgment will be required.

5. Who will be responsible for managing the organization's document base, or corporate memory? Ultimately, it will be desirable to have responsibility for data base administration and document base administration centralized under a single position. Currently, however, these functions may be divided between data processing/information systems and records management departments, which may or may not report upward to the same manager.

6. What types of searches will be formulated against an office's document base, and how is retrieval effectiveness related to the type of query? Buckland defined library usage in terms of four types of search (high document specificity and high information specificity, high document specificity and low information

specificity, low document specificity and high information specificity, low document specificity and low information specificity).(2) As the essence of retrieval is mapping an information requirement into a search strategy consisting of index terms, it is important to develop a similar taxonomy of information requirements in order to determine the level both of resources and effort that should be allocated to indexing so that the largest number of queries may be satisfied in a cost-effective manner.

Organization theory suggests that contextual variables related to uncertainty about an organization's external environment and the resulting demands placed on individual tasks will play a major role in determining the requirements for document-based information in a given organizational setting. Thompson, for example, stated that "uncertainty appears as the fundamental problem for complex organizations, and coping with uncertainty, the essence of the administrative process."(22, p.159) Subsequent research by Duncan identified two factors which contribute to overall perceptions of uncertainty. These factors are complexity, or the number of factors that must be considered in decision-making, and dynamism, or the degree to which decision factors change over time.(6)

One of the first individuals to link uncertainty to information use was Galbraith who defined uncertainty as the difference between the information that is required to complete a task, and the information on-hand.(9) Two subsequent field studies investigated the relationship between uncertainty and information usage. Holland et al. surveyed 384 scientists and engineers in a large government R&D organization and found a strong relationship existed between technical uncertainty and the choice of an information source. At high levels of uncertainty, scientists and engineers preferred richer, informal channels such as telephone conversations, and face-to-face discussions over documentary sources. They hypothesized, however, that their findings could in part be due to the fact that at high levels of uncertainty, the required information may not be available or easily transmittable through formal channels, and that the selection of an information channel may actually be forced in situations of high uncertainty.(10)

Blandin and Brown surveyed 70 top managers in electronics firms and wood products firms to "discover if systematic

differences existed in information search behavior for managers operating in environments characterized by differential levels of uncertainty".(1, p.115) They asked each manager to indicate the frequency of use for all sources, and the most important source used for acquiring external information from a list of four types of information sources (internal formal, internal formal, external formal and external informal). They found significant positive correlations between the level of uncertainty perceived by managers and 1) their reliance on external information sources, 2) their usage of informal sources of information, 3) the frequency of use of all information sources, and 4) the amount of time allocated to monitoring the environment. A positive, but not statistically significant correlation was found to exist between uncertainty and the frequency of use of formal sources of information (internal memos and reports, formal meetings and publications).(1)

Based on these findings, the following relationships are postulated with regard to document usage and vocabulary control in office information systems. First, as overall uncertainty about the environment increases, individuals will make greater use both of information external to the organization and of documents resulting from informal communication such as electronic messaging. Second, as dynamism increases, so will the use of documents resulting from informal communication. Third, as complexity increases, the types of documents used will increase and a more hetereogeneous vocabulary will be required in order to formulate effective search strategies across a number of topic areas. In summary, then, it is proposed that sub-units whose tasks are characterized by a relatively static, homogeneous environment will use a homogeneous set of documents which largely originate within the organization. Document retrieval can be accomplished using a highly controlled vocabulary. In heterogeneous, dynamic environments, a wide variety of documents will be used including documents originating outside of the organization, and retrieval will be accomplished using a much broader vocabulary. In addition, usage patterns in all settings will also be a function of informal communication roles (e.g. gatekeeper, liaison, opinion leader, cosmopolite or boundary-spanner: 18). For example, individuals who serve as a link between several sub-units in an organization, or between the organization and outsiders will employ a more diverse vocabulary than those who communicate largely within a single sub-unit. While it is expected that these propositions will hold true in both manual and automated settings, they should be verified empirically.

The implications for the design of office information systems are as follows. In settings characterized by a low degree of uncertainty (and a shared vocabulary), document originators could assign index terms with a great probability of successful subsequent retrieval by others because the same terms will be used by all to describe a particular phenomenon. In settings characterized by a high degree of uncertainty, there is apt to be less agreement on a common vocabulary and greater coordination across sub-units or tasks will be required. Further, it will be essential for systems in these environments to link users to sources of information outside of the organization. Finally, two caveats need to be stated. First, environments will vary across tasks within a single organization; it may be that multiple strategies for document processing will need to be adopted within a single organization. Second, great care must be taken in generalizing research results across organizations due to the impact of contextual factors (such as norms) on organizational information processing beyond the scope of those described above.

MANAGEMENT OF MANAGEMENT INFORMATION

Information management in the automated office includes a wide variety of functions ranging from records management to the management of an organization's internal models. This discussion will be limited in scope to the use of commercial information utilities.

The 1970's witnessed a spectacular growth in the number of commercial, interactive systems which provide ready access to external financial, marketing, legal, technological and general business information in the public domain. A number of these systems have been described in a previous paper.(20) These systems represent an important source of external information for an organization, and access to these systems on an individual basis is viewed as an important component of the automated office. However, a recent study found that managers rated these databases last of nine channels used to acquire external information both in terms of usage frequency and perceived accessibility.(5) There exist, therefore, major impediments to the use of these systems by individual managers. First, given the proliferation of systems, it may be difficult to identify the database or databases that are appropriate for a given task. Second, the command languages and indexing strategies vary from system to system; systems range from controlled vocabulary to free text. These differences represent a serious obstacle to successful query

negotiation once the appropriate database has been identified. Third, many of these systems are reference-based; they identify potentially relevant documents, but do not provide the full text. While in some cases, a copy of a document may be ordered from the information utility, it may often be difficult to acquire needed information on a timely basis.

Given these obstacles, it may well be that the best use of these systems from an organizational point of view is not made by individual managers. Currently, access to these systems is provided by an organization's library or other information units which is responsible both for performing searches and acquiring documents which the user considers to be potentially relevant. If the organization's library is a node in the automated office as well it should be, communication concerning a search request as well as the search results can be transmitted electronically with minimum inconvenience to the user.

Little is known about the ways that external information is both acquired and subsequently used to support decision-making.(17) Prior research related to organizational information processing, however, found that in a majority of cases, external information is acquired and filtered, and subsequently disseminated by a small number of key individuals known as gatekeepers or boundary spanners.(12) It is likely, therefore, that these individuals will make easy use of information utilities. Care should be taken to ensure that individuals who play this important informational role as well as other informal communication roles are connected electronically in the automated office.(18)

IMPLICATIONS FOR MIS RESEARCH AND EDUCATION

This paper has introduced issues associated with document-based information in the automated office which have not previously been widely addressed in the MIS literature. These issues have implications both for MIS research and education.

First, field studies in a variety of organizational settings should be conducted in order to develop a better understanding of the mechanics of document processing. The IBM study which led to the development of a prototype system might well serve as a model for future MIS research.(7) Such studies should investigate the volume and types of document usage, as well as the vocabulary associated with office documents. Descriptive studies of this type can provide the basis for the design of actual systems by relating information use to some of the organizational variables identified previously.

Next, prototype systems should be constructed and used for conducting both field and laboratory experiments in order to investigate the effect of a variety of indexing methods on system performance. Command languages will need to be tested both for flexibility and ease of use. The introduction of automated systems may result in different patterns of information use than existed in a manual environment, and these should be monitored. It will also be important to monitor retrieval effectiveness as the size of the document base available online increases. Cost benefit analyses related to synonym generation, and human versus machine indexing will need to be performed.

Third, both descriptive and experimental research related to the use of commercial information systems is also needed. Of particular importance is an understanding of the ways that these systems can become more accessible. Studies concerning the feasibility of developing a front-end processor so that an individual may utilize a single command language to access both intra-organizational and external sources of information should be conducted.

Research is also needed in order to understand how information is communicated within organizations and subsequently used to support decision-making. To be effective, the introduction of new technology should complement an organization's existing informal communication system. Future research should also consider the ways in which these informal roles change or evolve as a result of the increase in electronic communication and the subsequent decrease in face-to-face communication.

Finally, it will be necessary to broaden MIS curricula in order to address the issues associated with document processing. For MIS programs housed on a campus with a strong program in Library Information Science, joint programs or course offerings may be the most cost-effective solution. Information science programs have traditionally emphasized document processing related to bibliographic data; the typical product of such programs has sought employment in a library. Such individuals graduate with skills in indexing, thesaurus construction, query negotiation and actual use of commercial systems. The broadening

of information science course to include office as well as bibliographic documents would serve both library school students, by making them more marketable in a tight job market, and MIS students. A small number of information science programs are currently taking steps to do just that (the University of California, Berkeley, and the University of Denver), and The School of Information Studies at Syracuse University has developed a plan for offering a degree in Information Management jointly with the business school. At UCLA, a joint MLS-MBA degree is offered through the School of Library and Information Science. Where the possibility of joint offerings does not exist, MIS programs may seek to bridge the gap in their current curricula by recruiting faculty members from doctoral programs in information science.

The proliferation of office information systems may also necessitate a new type of knowledge among fledgling managers. Just as many business schools require a basic computer literacy of all their graduates, it may be in the future that a similar level of experience with commercial information utilities may be prerequisite to success in the business world. Service courses that are designed to acquaint an individual with the rudiments of database selection and online searching will need to be developed. The information age has arrived in the United States. Information management is now viewed as being as important as the management of financial resources in many organization. It is important that the MIS field meet the challenges provided by the emerging new breed of office information systems.

REFERENCES

1. Blandin, James S., and Brown, Warren B. Uncertainty and management's search for information. IEEE Transactions on Engineering Management, EM-24, 4, (1977), 114-119.

2. Buckland, Michael K. On types of search and the allocation of library resources. Journal of the American Society for Information Science, 30, 3, (1979), 143-147.

3. Bush, Vannevar. As we may think. Atlantic Monthly, 176, (July 1945), 101-108.

4. Culnan, Mary J. Alternative strategies for conducting online bibliographic searches in the social sciences. Unpublished (1978).

5. Culnan, Mary J. The impact of channel accessibility, boundary-spanning disposition and task complexity on external information usage in commercial settings. Presented at the National Meeting of the Academy of Management, (1980).

6. Duncan, Robert B. Characteristics of organizational environment and perceived environmental uncertainty. (ASQ), 17, (1972), 313-327.

7. Engel, G.H., Groppuso, J., Lowenstein, R.A., and Traub, W.B. An office communications system. IBM Systems Journal, 18, 3, (1979), 402-431.

8. Fahey, Liam, and King, William R. Environmental scanning for corporate planning. Business Horizons, 20, 4, (1977), 61-71.

9. Galbraith, Jay R. Designing complex Organizations. Addison-Wesley, Reading, MA., (1973).

10. Holland, Winford E., Stead, Bette Ann, and Leibrock, Robert C. Information channel/source selection as a correlate of technical uncertainty in a research and development organization. IEEE Transactions on Engineering Management, EM-23, 4, (1976), 163-7.

11. Luhn, H. P. A business intelligence system. IBM Journal of Research and Development, 2, (1958), 314-319.

12. McClure, Charles R. The information rich employee and information for decision-making: review and comments. Information Processing and Management, 14, (1978), 381-394.

13. McDonough, Adrian M. Study to determine the feasibility of using the syntax analysis computerized model by the Office of Management and Budget to help reduce or eliminate unnecessary and duplicative reporting in the federal government. Office of Management and Budget, (1980).

14. McNurlin, B.C. The automated office: part 1. EDP Analyzer, 16, 9, (1978), 1-13.

15. Meldman, Jeffrey A. Decision support systems for legal research and analysis. MIT Sloan School of Management Working Paper, CISR #38, (1979).

16. Morgan, Howard L., and Root, David J. A concept of corporate memory. Proceedings of the NYU Symposium of Office Automation, (May 1979).

17. O'Reilly, Charles A., and Anderson, John C. Organizational communication and decision making: laboratory results versus actual organizational settings. Presented at the XXIV International Meeting of the Institute of Management Sciences, (June 1979).

18. Rogers, Everett M., and Agarwala-Rogers, Rekha. Communication in Organizations. Free Press, N.Y., (1976).

19. Schwartz, Ronald, Fortune, Joan, and Horwich, Julian. AMANDA: A computerized document management system. MIS Quarterly, 4, 3, (1980), 41-49.

20. Swanson, E. Burton and Culnan, Mary J. Document-based systems for management planning and control: a classification, survey and assessment. MIS Quarterly, 2, 4, (1978), 31-46.

21. Teichman, Ronald J. Development of information systems activities and interfaces to support top management functions. Unpublished Ph.D. dissertation, Northwestern University, (1977).

22. Thompson, James D. Organizations in action. McGraw-Hill, N.Y., (1967).

23. Wohl, Amy D. Replacing the pad and pencil. Datamation, 26, 6, (June 1980), 169-176.

THE KNOWLEDGE AND SKILL REQUIREMENTS FOR THE DOCTORATE IN MIS

GORDON B. DAVIS

The University of Minnesota

1. INTRODUCTION

A doctorate in MIS is primarily to prepare for an MIS education career, although some persons with MIS doctorates go into consulting or into industry. There are (in 1980) probably between 200 and 300 persons with MIS or essentially MIS doctorates. The current annual rate of graduation of MIS and near-MIS doctorates is estimated at 50 per year. The MIS doctorate is not a uniform product; there is substantial diversity based on the college, the faculty, and individual program design.

The body of knowledge for a doctorate in MIS is defined by four methods: derivation from MIS subject content of MIS bachelors/masters degrees, a derivation from a model of MIS research, an examination of the intersection of MIS with other fields or study, and a reading list. This MIS domain is large and diverse which means that an MIS doctorate cannot be expected to have uniform depth of knowledge across all knowledge topics that are included in the domain.

The research skill requirements for receiving the MIS doctorate are similar to those expected of most doctorates in business or social science. However, a person having an MIS doctorate should also be familiar with MIS research frameworks, the application of research strategies to MIS topics, and the management of knowledge work. Doctoral-level workshops and seminars are important in developing depth knowledge in MIS concepts, theory, structure, and research.

There are about 30 universities that have indicated a capability to offer a doctorate with a major in management information systems. Not all of these universities provide the same level of support, but they provide a starting list for a person considering an information systems doctorate.

The purpose of this paper is to provide a basis for discussion and exploration. (This paper uses some material from an article by Gordon B. Davis, "The Education of Information System Educators," _Interface_, Fall, 1979.) It is not prescriptive in the sense that a single "answer" is developed. The diversity of doctoral programs makes it undesirable to define a single normative set of skill and knowledge requirements for an MIS.

2. ASSUMPTIONS ABOUT THE PURPOSE OF A DOCTORATE IN MIS

The doctoral degree in management information systems should prepare a person to:

1. teach

2. do research, and

3. interact with the information systems community.

Some who obtain the doctorate in information systems may choose to be consultants or work in industry; however, the major purposes of the doctorate in the information systems area (as in other subject areas related to management and organization) are teaching and research. Table 1 illustrates this point for Minnesota Ph.D.'s in MIS. Partial returns from a survey of Ph.D. granting schools show approximately the same results.

TABLE 1

Minnesota MIS Ph.D.s, 1971-79

Current Occupation	Number	Percent
College Professor in MIS	16	70
Consultant	4	17
Industry Executive	3	13
	23	100

This paper will therefore focus on the doctorate as a preparation for a career as a college professor. It does not address the issue of vocational-technical or community college instructors who may need practical experience, but do not need a doctorate.

174