Improving corporate information services in an automated word-processing network

by HENRY L. MAYFIELD
Shell Oil Company
Houston, Texas

ABSTRACT

The flow of corporate information is destined to change considerably with the arrival of word-processing technology and the expansion of computer communication networks. Computer programs will extract indexing information from documents prepared in word-processing clusters and pass the data to a central computer complex for filing. By querying that database, employees will gain access to documentation corporate-wide. This paper follows the steps in a document’s life cycle. A methodology is examined by which corporate information service departments may plan for, and ultimately provide, the benefits of enhanced information access.

INTRODUCTION

Attention has recently focused on the reaction of corporate management and office support staff to revolutionary change brought with office automation. Many organizations, through distributed word-processing centers, document improved clerical productivity. Several experiments in electronic mail cite speedier information flow as having a positive impact on high-level decision-making. Largely ignored, until now, is the methodology by which the corporate information service department (CISD) will enable almost all corporate employees to derive immediate benefit from the new technology.

The potential beneficiary of improved information services is anyone who comes in contact with textual documents. The known benefits of word-processing are seen in the composition cycle of letters and reports. With the use of these services, corporations are better equipped to produce paper in what has been termed the "technical information explosion era." Present world-wide information growth rates of 12-13 percent are predicted to be four to seven times as large in the mid-1980's. Employees flooded by unstructured information will find they are ill-equipped for minor job crises unless new mechanisms are devised to aid them in the document reaction cycle. Standards and services, developed by CISDs, could act as one such positive mechanism.

Ultimately, employees will have access to computer reports indexing documents authored by themselves or their co-workers. (This paper does not explore information access to extra-corporate literature.)

Document retrieval from a locally available micrographics archive will then take only a few minutes. Supervisors, higher management, and other authorized personnel will dial into a central computer database and retrieve any (externally transmitted) documents by typing in keywords describing their content. Some documents will be transmitted via electronics. Others which formerly were reproduced in great quantity prior to shipping will arrive in an envelope on microform and be printed at the receiver’s site. When appropriate, information specialists will provide personalized assistance to the information-seeker.

Before looking at how all of this may be achieved, a review of present systems—modified by word-processing/computer network technology—is in order.

INFORMATION FLOW IN THE NETWORK

The components of an automated word-processing network, including communications linking terminals and/or computers at the network nodes, have passed through several phases:

Phase 1—Communications link remote-job-entry or timesharing terminals to a physically distant computer complex.
Automatic text-editing systems permit operators to key-enter and edit documents using low-speed typewriter/CRT devices. Documents are filed on disk storage at the central complex.

**Phase 2**—Installation of “intelligent” terminals at the satellite nodes permit documents to be stored locally on cassettes or diskettes.

**Phase 3**—By replacing intelligent terminals with multi-terminal minicomputers, the enlarged network serves a greater number of operators. These low-cost shared logic systems placed within a local word-processing cluster replace the function of the distant central system of Phase 1.

Phase 3 networks can be utilized in several ways. Most user composition activity takes place at the network's satellite nodes. Documents are entered, edited, and printed within the cluster. When the network is used for distribution, a document may be transmitted electronically via communication links to other network nodes for printing at the addressee's cluster. When it is used as an electronic message center, the addressees are notified at their on-line terminals of the newly-arrived document. They may check the sender, date, and subject of the documents before printing or destroying selected ones. Replies may be composed and distributed electronically.

A method for document filing, storage, and retrieval to aid the “reaction cycle” spoken of earlier is being evaluated by researchers at the Massachusetts Institute of Technology. Their DMS Message System takes document messages and transmits them to other points at a satellite node or to other computers attached to the ARPANET communications network. The ARPANET, begun in the 1960's, is an experimental network linking over 100 private and public computer/research installations from Hawaii to Europe. Users are able to lower their computer operating expense by sharing computer resources, e.g., central processing units, mass storage devices, plotters, and microform systems.

In preparing the documents for filing, messages are broken down into fields, such as subject, date, text, action-to, carbon-copy, etc. The assigned field-values are used to index and store/retrieve documents to/from a database. How DMS would economically interface with a variety of vendors' computer hardware/software in a corporate network is not clear at this time. Conceptually, features of DMS are certainly among those that could enhance corporate information services.

**ENHANCING SERVICES IN THE NETWORK**

There are two immediate goals that CISDs should seek in a Phase 3 network:

1. Specification and enforcement of corporate documentation standards. When satellite word processing clusters use common report-numbering, author/name, and date-entry formats (among other documentation standards), the problem of computer indexing becomes a manageable function.

2. Word-processing clusters able to communicate with each other and with a central computer complex. Local word-processing terminals unable to communicate with terminals elsewhere in the network fail to maximize their usefulness in an electronic message environment.

These goals, together with a corporate records management program, put CISDs in a position to meet the present demand for information at a lower cost, while increasing overall service levels.

In reality, fulfilling the two goals may parallel the actual services improvement, with each activity reinforcing the other. In analyzing how this may evolve, composition, distribution, and reaction cycles of a document will be examined in detail. Emphasis will be on a methodology which can be integrated into any Phase 3 network meeting the two objectives cited above, regardless of vendor software/hardware utilized.

**Composition**

“Information mapping” is a method for defining and enforcing corporate documentation standards. Originally used in training materials on technical subjects, mapping structures textual writing by dividing the page with horizontal and vertical lines to form logical blocks. Only information of one functional type exists in a given block. The concept, while extendable to almost the whole of corporate documentation textual in nature, will be illustrated in the context of internal and external correspondence.

The typical business letter or memo may be divided into logical blocks, as in Figure 2. It should be noted that the text-block could relate to abstracts for other types of documents.

Within each of these blocks, a typist enters words

![Figure 2—Information map](From the collection of the Computer History Museum (www.computerhistory.org))
fulfilling a unique function—a type of information map. Chances are that not too far from the typist’s keyboard is a secretarial handbook outlining the function and format for each block: specifications for vertical and horizontal spacing; address format and punctuation; in internal correspondence, when to use top-down or bottom-up departmental hierarchical TO, FROM identification; layout for dictator, typist, enclosures, copies, filing, references, attachments.

When correspondence is prepared on conventional typewriters, strict standards are often compromised. When using “intelligent terminals” programmed to validate operator entries, the skill in and practice of documentation standards are improved. At Dun & Bradstreet, Inc., an operator types a continuous string of characters prefixed by the logical block into which they are to be stored for printing. When applied to information mapped correspondence, the typist may enter:

:DATE: December 1, 1976
:TO: A. R. Jones President, Software Inc.
2100 Broadway
New York, NY 10000

:GREETING: Dear Art:
:TEXT: Further to your letter of . . . .

. . .

and early resolution of the problem.
:FROM: Sincerely, T. R. Smith
:OTHER: DRJ/SWC

This type of data entry has been found superior to traditional CRT template orientation. Standards of 12,000 strokes per hour (equivalent to a typing speed of 40 wpm) have been established. Blocks of a document may be entered in any order and immediately validated for content, format (punctuation, indentation, spelling), and block length. The program which performs these activities may also establish page-positioning for display. Dun & Bradstreet has found that a new terminal operator can reach full productivity in one-quarter of the time that would be required in using a manual system.

Indexing/keywording

The most important aspect of standards established and enforced via information mapping comes to bear in computer indexing of documents.

With text-editor/data entry systems operating on a multi-terminal minicomputer, the indexing of material in the cluster is now feasible. Even when equipment manufacturers provide no generalized software specifically supporting data entry based upon information mapping, programs tailored to corporate specifications are installable. Indexing a document, in which physical position on the page defines the context in which keywords and phrases are used, is a straightforward approach to an otherwise difficult problem.

At Shell, a computer program has been written to process correspondence as mapped in Figure 2. Several activities are carried out by that program:

1. Non-text blocks are validated for content and format.
2. Non-text blocks are indexed (as multi-valued fields, when required).
3. As an alternative to keying all significant single words, keywords and phrases specified by the document’s author are extracted from the text blocks.

The distinction at this point between “indexing” and “keywording” is important. In indexing correspondence, for example, an individual name (“JONES, T”) appearing in the TO or FROM or COPIES blocks is an index field value. The context in which “JONES, T” is used is clear-cut, both in how a computer would store the index field-value and how an information-seeker would query the computer database:

FIND TO IS ‘JONES, T’

A document with multiple authors would be assigned multiple index field-values:

FIND AUTHORS ARE ‘JONES, T’ AND ‘SMITH, A’

On the other hand, ‘JONES T’ appearing as a keyword in the documents’ text-block is subject to varying degrees of usefulness in retrieval:

FIND ‘JONES, T’

might retrieve documents in which the name was only mentioned incidently or miss documents where ‘T JONES’ appeared in the text-block. Clearly, terms which appear in document text do not always provide an adequate basis for retrieval, even when specified by the document’s author. Automatic extraction of all non-stop keywords, like speci-
fied keywording, may at times produce too few or too many terms, causing a wide variability in the completeness and relevance of documents retrieved. Thus, other keywording is advised.

Extra keywords which enrich or clarify the subject content of documents may be assigned by a local information specialist from a controlled corporate thesaurus. Entered as a special information-map block, the keyword contents would not be routinely displayed in the distributed document.

**Distribution**

Once a document's composition cycle and indexing/keypunching based upon information mapping concepts are completed, the delivery may be carried out in several ways. The document itself may be printed within the word-processing cluster and transmitted via regular mail. For speedier transmission to a physically distant addressee, the network may be used to transmit the document in a fraction of the time it took for initial entry (based upon an average 2400-characters-per-second transmission speed versus 40 WPM entry rate).

Indexing and keyword information may be held in the database of the local cluster for security considerations or for documents of narrow interest. Otherwise, indices, keywords, and the document itself are transmitted to the central computer for processing by the CISD.

**Storage**

Textual documents historically have arrived at CISDs in printed form. The effort in manually indexing, filing, and micro-filming documents of diversified formats is a task many CISDs are already hard-pressed to handle. With the projected increase in printed matter arriving at CISDs, staff-time now given to information seekers may be shifted to meet the increased indexing and filing burden. The completeness and relevance of information retrieved is bound to degrade unless the two objectives cited earlier are sought.

In an enhanced services environment, computers will perform many of the indexing, filing, and retrieval functions that earlier required human manual and intellectual effort. The information specialist will coordinate the receipt of electronic messages from all nodes of the corporate network that have been directed to the central computer complex. Electronic transmission of documents from one node to another need not pass through the central node. However, if the document sender or recipient wish to participate in the new services provided by CISD, the indices/keywords and, optionally, the document itself must be enqueued for processing by the CISD specialist. Filing and retrieval services provided by CISD include:

1. Processing (computer) index/keyword messages. Validate format and control posting in corporate database.

   Maintenance and distribution of corporate thesaurus for keyword enrichment.

2. Collecting documents for archival tape storage and output to film.

3. Preparing and distributing micrographic files to designated word-processing clusters, with their accompanying computer generated subject-indices.

4. Maintaining employee interest profiles. Matching profiles to updated database. Distribution of relevant information via the network or regular mail.

**Retrieval**

With the document life cycle described above, a mechanism now exists to help anyone in the document reaction cycle.

The recipient or author of a document mailed electronically need only retain a printed copy until the arrival of the next set of micrographic files in the departmental word-processing cluster. Archival retrieval may always take place using the computer-generated subject index and film readers located in the cluster. If extra copies of a document are needed, a film-to-paper copier is available.

For documents not authored or mailed to the departmental cluster, authorized employees may interrogate the CISD maintained corporate database on the central computer. Queries are entered at a low-speed terminal, transmitted over the network to computer programs interacting with the database, and the results returned to the user.

A variety of traditional retrieval tools are available to aid the on-line information seeker. In formulating a query which will be complete and precise, a user may selectively display at his terminal:

1. An alphabetical thesaurus listing of index fields (TO, FROM, SUBJECT, TITLES, AUTHORS) and keywords with a count indicating the frequency with which the terms appear in the database.

2. Vocabulary analysis of terms, indicating possible alternate terms, e.g.,

   —SEE ALSO'S—
   Data Management—See File Management

   —GENERIC RELATIONSHIPS—
   Information Systems Manager—
   BROADER TERM—Center General Manager
   NARROWER—Information Design Manager
   Information Support Manager

Retrieval using index/keyword techniques discussed earlier, together with aids for query formulation, has been measured and the performance judged competitive with pre-coordinate, strictly controlled keywording approaches. If further retrieval assistance is required, a user's query may be processed by a CISD specialist at the central complex and the results displayed on the user terminal.
within the cluster. Together, they are able to obtain sufficient information to learn which documents are relevant to the information seeker's interests and where the entire document may be retrieved. For locally authored/received documents, the film reader may once again be consulted. Another alternative is to interface a CRT to a micrographic retrieval device, which in turn is linked to a database index. When the computer responds to a user query, the resultant micro-image may be transmitted directly to the microfilm reader and the image displayed automatically. Otherwise, there are several alternatives for delivery of the document set:

1. Transmission over the network for high-speed printing within the cluster.
2. Normal mailing of batch print-out from the central computer.

OTHER CONSIDERATIONS

As the services described above are put into place, the benefits which follow may be measured by more than just the availability and immediacy of information.

Word-processing is estimated to yield increases in document productivity of 100 to 500 per cent. With the saving, an initial capital equipment expenditure of $15,000 to $20,000 per cluster operator station is expected to be paid out in two or three years.

The one-time expense of manual filing, mailing, and retrieval of a 750-character document is estimated at $.59. To store the document prepared via word-processing for on-line retrieval would run approximately $.004 cents per month. A retrieval using one of several popular storage and retrieval systems would run $.05 to $.25, depending on a variety of factors.

The expense of present information services versus enhanced services in a word-processing network environment may justify the implementation of the latter.

Other benefits are less obvious. In order for the word-processing network to function properly, there exists within it a directory of corporate employees. Each is identified by:

1. Employee number
2. Name/Title
3. Cluster assignment
4. Hierarchical department/function/role designation
5. Network security authorization
6. Telephone number

With the directory, documents may be related to authors/recipients with employee numbers alone. The formatted report, depending on its type, will substitute the appropriate employee text. Tags grouping different employee collections or hierarchical department designations may be built and stored to aid in the automatic preparation of routing information. Combinations of employee number, departments, and security authorizations will also be used to permit only qualified individuals to interrogate cross-sections of the CISD document database. The directory may serve other purposes as well.

As an interface with the corporate personnel database, the inventorying of employee skills, training, job history; the preparation and distribution of payroll are aided. The directory also acts as a convenient source for compiling organization charts and telephone directories.

Improved documentation standards and services, if adopted within the cluster, mean that personal filing practices may be impacted. The employee and a cluster information specialist will enter keyword/indexing information into the local database. Film and paper files are maintained by the technologist, leaving more time for the employee to pursue normal job responsibilities.

CONCLUSION

It is said that 5-10 years will elapse before the declining hardware costs will make widespread corporate word-processing clusters commonplace. In that time-frame, "... the cost of communications systems will drop so far that all kinds of applications which seem exotic today are going to become feasible and attractive," Arthur D. Little, Inc., has reported that by the 1980's, decreases of 50 percent for central processing, 66 percent for communications, and 80 percent for terminals are likely. Costs for storing a page of information will decrease by a factor of 30 when paper is replaced by other media.

In the interim, lack of planning could result in divergent objectives and disparate systems implemented throughout the corporation. The potential for waste and duplication of effort is very real in parochial approaches to corporate word-processing.

Information service departments should move to:

1. Establish documentation standards for use in word-processing clusters to facilitate automatic indexing;
2. Cast a frame-work for decision-making in cluster implementation and operational procedures;
3. Coordinate planning in anticipation of providing full-fledged services in a Phase 3 word-processing network.

ACKNOWLEDGMENT

The comments and support of H. D. Baade, M. K. Johnson, and J. M. Wainright were invaluable in the drafting of this paper.

REFERENCES

Does it Fit?,” Massachusetts Institute of Technology, Laboratory for Computer Science.


