

having various numbers of coordinate index terms and subclasses and for search sequences. Preliminary results indicate that coordinate index terms will considerably reduce retrieval by the classification schedule alone and will still provide representative desired documents.

The next objective was to reveal specific needs, if any, for modifying and correcting the system before changes become too costly. Some modifications of the original term lists were made as an indirect result of the evaluation program in the preparation of the sample files. The expected number and distribution of documents retrieved suggests a need to investigate further methods for physically retrieving documents once they are identified.

Another objective of the evaluation was to investigate indexing accuracy of Examiners, Classifiers, and Analysts as potential indexers. The indexing experiment showed that input errors involving coordinate index terms should not be an appreciable source of search failures if satisfactory input standards are maintained. The four indexers chosen for preparation of the file all had satisfactory indexing accuracy.

The fourth objective was to ensure that the system will perform satisfactorily from the standpoint of input accuracy. The system appears to be satisfactory in that few relevant documents should be missed either from indexing errors or errors in searching if the proper search strategy is used.

The final objective was to provide a preliminary investigation of search strategies. The study yielded good information concerning potential alternative search strategies. However, further studies of search strategies are anticipated in an operational search experiment to further

improve search results and to train Examiners in the use of the system.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance and cooperation of H. S. Vincent, project coordinator for the design and development of the combined search system, and to thank especially A. J. Cook, F. J. Cullen, D. J. Funderburk, R. L. Huff, A. Mazel, D. B. Moyer, and S. Rosen who participated in the indexing experiment.

LITERATURE CITED

- (1) King, D. W., *J. Chem. Doc.* **5**, 96 (1965).
- (2) Vincent, H. S., "A Combined Manual and Mechanized Search System," Proceedings of the Fifth Annual ICIREPAT Meeting, in press.
- (3) McDonnell, P. M., and Vincent, H. S., "A Combined Manual-Mechanical Search System for Patents Relating to Medicines, Poisons and Cosmetics—Present Status," Paper presented at the Sixth Annual ICIREPAT Meeting, The Hague, October, 1966.
- (4) The term "relevant" is used in the sense of being useful to the Examiner. Relevant documents are those cited in source patents. Mere citation, however, cannot be construed as an indication of what one commonly considers relevancy. An issued patent contains a list of all references cited during prosecution of the case for any reason, such as requirement for restriction, rejection of claims later withdrawn, etc. Furthermore, no cited reference fully anticipates the claimed subject matter of the source documents.

Designing an Author-Based Correspondence Information System*

HERMAN SKOLNIK and MARION H. PAYSON
Research Center, Hercules Incorporated, Wilmington, Delaware

Received May 16, 1966

The problems associated with a centralized correspondence information system in a research and development environment are examined. These problems relate to receipt and control, retention periods, circulation, filing, and retrieval. Various mechanisms for solving these problems are considered, and a correspondence information system based on filing alphabetically by author is described.

In many research and development environments, technical correspondence is a storehouse of information almost equal in importance to the report literature and to the library. Indeed, these three categories of document

holdings are considered so important in some research and development centers that they are maintained as three separate and distinct document areas. The needs of scientists for, and the nature of, these three categories are such that their separation is often well advised. From the viewpoint of protection and control, the philosophy and policies invoked for reports and correspondence are completely different from those for a library.

*Presented before the Division of Chemical Literature, 152nd National Meeting of the American Chemical Society, New York, N. Y., Sept. 13, 1966.

This paper examines the problems associated with a centralized correspondence information system in a research and development environment, and discusses various mechanisms for solving these problems.

A correspondence information system is anything but simple, yet no other document form is taken so much for granted. Correspondence is the most common of all document forms. Everybody writes and receives letters and practically everybody has his own personal correspondence files, which, more often than not, are a secretarial province. It is not surprising that the education of secretaries includes a course on the maintenance of letter files. It is surprising, on the other hand, that Library Science and Information Science curriculums ignore the problem almost completely. Consequently, most correspondence files have been organized and are operated by secretaries on the principles of general office practice.

Whereas general office practice may be a satisfactory basis for the operation of office correspondence files, particularly when operated by an intelligent secretary, centralized technical correspondence files must be designed from the communication and information system viewpoints. Important factors in the design are:

1. Receipt and control of correspondence.
2. Retention schedules.
3. Circulation of correspondence.
4. Filing system.
5. Retrieval system.

RECEIPT AND CONTROL

Receipt of correspondence must be considered from two aspects: outgoing and incoming mail. The best means of assuring that copies of all outgoing mail go into the correspondence information system is to train typists and secretaries always to make a carbon copy for the correspondence files. Having these file copies always of the same color in the carbon copy set is a practice that practically guarantees that the copies will get into the system. Multilith copies, on the other hand, cannot be easily and economically produced on a color set, particularly for intracompany communications. In this circumstance, the duplication stencil or master must show the correspondence files as receiving a copy of the document; or typists and secretaries must be trained to send a copy of all duplicated correspondence to the files; or arrangements must be made for the duplication operation to send a copy of everything it produces to the files. The last is probably the most foolproof arrangement.

Incoming mail is another matter. At the Hercules Research Center, it is received and opened, if not marked personal, by the mail room. If the correspondence system is good, all incoming technical mail will be delivered immediately to the correspondence files for processing and circulation. This is the critical point in many centralized correspondence files. Although we could enumerate several possible shortcomings of correspondence files, the most apparent shortcoming is the failure to realize that automatic receipt of all incoming technical mail by the correspondence files is essential for proper control and communication. If the correspondence system is slow and

inefficient, and has not gained the confidence of the technical staff, we would recommend microfilming or photocopying incoming mail as received rather than delaying its circulation to the addressee and others.

If the correspondence information system is good, however, incoming mail will be processed and in circulation before the next incoming mail delivery. This requirement can be met in a properly designed system. In a large sense, the ability to meet this requirement is a measure of the efficiency of the system.

The processing steps are the control operations for acquisition, retention, circulation, storage, and retrieval (Figures 1 and 2). Actually the first step occurs in the mail room, where every letter is date and time stamped. Every letter received, however, is not necessarily retained in the correspondence files. As for any document file, correspondence needs to be defined in terms of what it does and does not include. For example, mail relating to the functions of purchasing, personnel, and administration probably is better excluded from a technical correspondence file. These three categories of mail in a research and development environment are usually of such a nature as to merit their own separate operations. Thus, one of the functions of centralized technical correspondence files is to separate the mail for each of the other files, retaining only that which is related to and pertinent to all technical aspects of the research and development programs. Technical correspondence also includes patent cases and letters of historical interest to the company.

The letters which are to be retained by the centralized technical correspondence files should be clearly marked for return to the files. An illustration of a stamp for marking correspondence is shown in Figure 3.

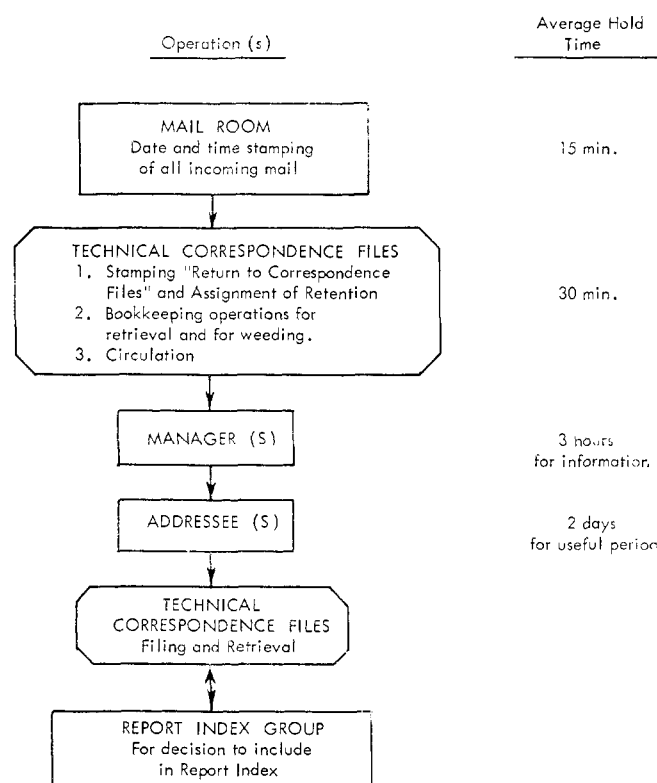


Figure 1. Flow sheet for incoming correspondence.

RETENTION SCHEDULES

Retention schedules for technical correspondence need to be in harmony with corporate policies as well as with the technical needs of the environment. Corporate policies for retention of documents are usually established by a corporation's treasurer, secretary, or legal director to ensure uniformity throughout the corporation in business and legal matters. Retention schedules outside of the defined areas of these policies are generally assigned a retention of "useful period."

Letters of a general technical, business nature usually have a 10-year retention period. Letters containing information of value to the obtaining of patents have a retention period of at least 20 years, and sometimes of "permanent." A permanent retention period is also usual for letters concerned with corporate policies or letters which contain information of historical interest to the company.

Although a research and development environment may have an efficient technical report system, scientists may choose to report their laboratory observations and conclusions in a letter. This decision is usually prompted by a need to communicate which is outside of the established channels. Letters of this nature are essentially equal to a technical report and consequently should be considered for permanent retention. At the Hercules Research Center, these letters are routinely circulated to the Report Index Group for the possibility of including them in the Report Index. (The letters are filed, however, by the Correspondence Files, not by the Report Files.) If the Report Index Group indexes and abstracts these letters for the Report Index, the letters are marked "P" for permanent retention.

A one-year retention period might be considered for letters on meeting and visit dates, requests for literature, and personnel involvement with scientific organizations and publications. It is an advantage to have the retention period established as soon as possible, certainly before the letter is filed. Our Correspondence Files stamps letters as shown in Figure 3 and immediately indicates the retention period with a numeral or "P" in the space provided. Letters of questionable value are stamped with an additional stamp as shown above the dotted line in Figure 4. The major advantage of the double stamping is the involvement of the addressee in deciding the retention period in terms of his needs.

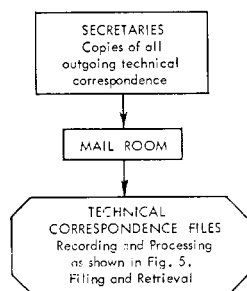


Figure 2. Flow sheet for outgoing correspondence.

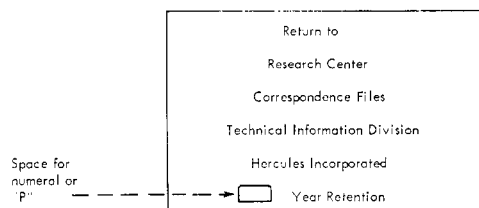


Figure 3. Rubber stamp for marking correspondence. P is for permanent retention.

CIRCULATION

Circulation of intracompany and outgoing mail is a relatively minor matter as this can be controlled, generally when the letter is typed, with names of all those receiving copies. Nevertheless, even here, the correspondence information system can serve an important function in routing a letter to those with a need to know but overlooked by the writer.

On the other hand, circulation of incoming mail can be a problem of conflicting communication needs. In the first place, incoming mail is usually a single copy. Although photoduplication is an easy and relatively inexpensive process, it becomes a time-consuming and costly operation if applied to many thousands of letters per year. An over-all communications policy based on need-to-know and prompt attention can reduce photoduplication of mail to cases of high urgency only.

The addressee of an incoming letter does not necessarily have the greatest need for the information nor does he necessarily know who may have the greatest need. In a research and development environment, where an incoming letter might affect work assignments or operational policies, the addressee's supervisor or manager and the director have a greater need to know the information. It is relatively easy to define these areas to the satisfaction of the technical staff, particularly if the supervisors, managers, and the director give this mail prompt attention. If the supervisory staff do not give this mail prompt attention, they will have a disgruntled technical staff on their hands. But then the correspondence files supervisor will be the first to hear the complaints and, accordingly, will inform those who can initiate steps to solve the problem.

FILING SYSTEM

There is no single best filing system for correspondence. Many factors must be considered in selecting it. Some important ones are:

1. Size and complexity of the holdings.
2. Allotment of space for the files.
3. Education and training of the files staff.
4. Number of potential users.
5. Variety of ways and most likely way the users request letters.
6. Adequacy of the retrieval system.

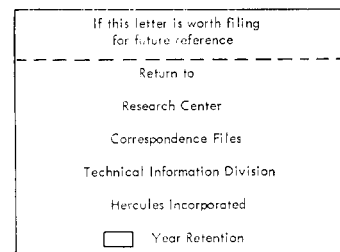


Figure 4. Double stamp for marking correspondence when retention period is not obvious.

Let us consider several filing systems and how the choice may be affected by the above factors.

According to the literature on correspondence files, the overwhelming choice is to file letters by a classification system. There are many reasons for this situation. Secretaries are taught to file this way. Librarians and some chemists tend to think this way. Individual files, and small office files, generally require that letters be filed in accordance with some classification scheme set up by the boss or by his "first" secretary.

The next most favorite scheme is to file correspondence by subjects.

Both classification and subject systems can be excellent for correspondence files. In a technical correspondence file that processes many thousands of letters, the information can be relatively highly technical and complex, and most likely would require the presence of a knowledgeable technical person for the assignment of classification or subject terms for each letter. Furthermore, classification and subject schemes require that there be as many copies of a letter as there are classes and subjects. If space is restricted, duplicate filing of letters can become a problem. It certainly adds to the cost and to the workload.

Other filing systems are: by research and development project number; by accession number; by author. The disadvantage of the project number is most apparent with incoming mail. The accession number system is generally used in connection with a concept correlation retrieval system.

Since the Hercules Research Center Correspondence Files was set up in 1927, it has evolved through the following filing systems: project number, subject, classification, and, since 1955, author. The evolution occurred in harmony with the growth of the company and with the expansion and maturation of the research and development programs and staff. For example, in 1927 the company manufactured and sold several dozen chemical products; the products today exceed 1000. In 1927, there were several Ph.D.'s on the research staff; today, there are hundreds. In 1927, we processed a few thousand letters; in 1965, we processed over 40,000.

In 1954, the last year in which correspondence was filed by a classification system, the Correspondence Files staff consisted of one chemist and two file clerks (who processed 35,000 letters). In 1965, a staff of 1.5 file clerks processed over 40,000 letters and thousands of requests for letters.

A six-month study was carried out in 1954 to establish the various needs of the technical staff for correspondence—*i.e.*, a specific letter versus all letters under a generic or specific subject or class. Over 80% of the requests were for specific letters by a Hercules author or from an outside company, and a large fraction of the generic subject or class requests was related to the author or authors of the information. Furthermore, there were more requests for all letters by a given author or from a given company than for all letters on a given subject.

Consequently, the filing system was changed to author (and outside company) filing in 1955. Since then, we have continued to study the effects of author filing on the processing operations, service to the users, and retrieval of information.

Advantages of Author Filing. Author filing, as already indicated, is a primary point of retrieval at Hercules. An important result is that 80% of the requests are taken care of in terms of a few seconds per letter. The remaining 20% are handled by means of the retrieval system which will be discussed below.

Whereas the subject and classification system files required a chemist for supervision and file clerks for book-keeping (assigning subject or classes, and checking letters in and out) and filing operations, author filing operations involve only (and fewer) file clerks. This staff reduction is possible for the following reasons:

1. No need to write and check the classification or subject on the letter.
2. No need to duplicate the letter or prepare dummy sheets, as author filing requires that only one copy be filed.
3. Because there is only one place for the filing of a letter in an author file, filing space and time and retrieval time are the least possible.
4. No need for a classification or subject schedule, an expensive and time-consuming operation.
5. Author filing, in our experience, requires fewer file folders and changes in the file folders from year to year than subject or classification filing. A concomitant advantage in this operation is the ease in separating recently dated letters, such as the current year's, which have the highest request rate, from the rest of the files. This separation is easy in author filing, but extremely time-consuming in classification and subject filing.
6. Author filing of correspondence can be taught in a few minutes, whereas classification and subject filing require many teaching hours and considerable experience.

RETRIEVAL SYSTEM

If the correspondence files are to be run by file clerks, the retrieval system must be designed from their viewpoint as well as from that of the potential users. This is less difficult than appears at first sight, providing there is a good understanding of the needs of the users. The highly technical (reportlike) letters, which are relatively few in number, obviously should be indexed by chemists. We take care of this category by processing such letters through the Report Index Group for incorporation in the Report Index. The Report Index is available to the correspondence clerks and to all potential users of the correspondence files. Consequently, the retrieval system for the correspondence files can be designed at a somewhat lower technical level.

Information in letters to or about an outside company is an important retrieval viewpoint for the users. A card file (one card per company on which are posted date, author, and keywords) plus the company folder or folders give a complete to-from-about record. Figure 5 illustrates a company card.

Although the most frequent request is satisfied by author filing, and the next most frequent by the company card file, about 5% of the requests require the use of an index for retrieval. This does not mean, however, that all letters need to be indexed. On the contrary, a study of the informational content of the letters revealed that no more than half required retrieval viewpoints beyond the author

ABC COMPANY		
3/17/66 → 10	John Doe	S - Resin
3/20/66 → 1	Jack Smith	visit 4/27

Figure 5. Company card.

Arrow slanted upward indicates a letter to the subject company.

No arrow indicates that a letter is about the subject company.

and company. The remainder of the letters required some indexing from the product, application, process, and apparatus or equipment viewpoints and whether a sample preparation and shipment were involved. These viewpoints are relatively obvious in the letters and are not difficult for the file clerks to enter in the correspondence index.

The correspondence index consists of looseleaf notebooks in subject alphabetical order, one subject per page for posting letters by author, date, and a few keywords. In the event that a specific author's correspondence dominates a subject, a situation which we have discovered is relatively frequent, a cross reference to the author is entered on the subject page.

When a correspondence information system is responsive to the needs of the users and gives efficient and effective service, the users tend to communicate their changing needs and to suggest subject areas for these needs. Indeed, an increasing use of the correspondence information system and a concomitant increasing feedback from the users have been noticeable effects of changing to author filing.

SUPPLEMENTARY INFORMATION

Bookkeeping operations are essential in all correspondence files, yet must be kept to a necessary minimum. We have found that author filing requires fewer bookkeeping operations because there is only one place for each letter. A log sheet for each author (or company) is an excellent mechanism for knowing whether or not a letter was ever received and to whom the letter has been circulated. The log sheet is attached to the file folder as a permanent record. We have found the log sheet to be helpful in searches, particularly when a

requester knows the author but is uncertain as to the date.

Weeding of letters in accordance with retention schedules is set up before the letters are filed. A red "flag" tape is attached to the upper right of the 11-inch side of letters having a one-year retention and on the right of the 8-inch side for letters having a 20-year or permanent retention. (The tapes are Brady Self-Sticking Perma-Code wire markers, obtainable from W. H. Brady Co., 722 West Glendale Avenue, Milwaukee, Wis. 53209.) As the tagged letters are obvious in a folder, they are readily and quickly separated in the yearly weeding operation. Figure 6 illustrates the two kinds of tagging.

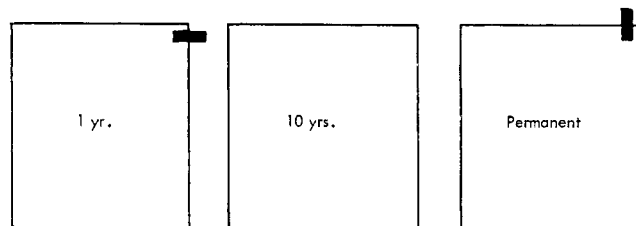


Figure 6. Tagged letters for weeding

The filing of correspondence in the Hercules Research Center Correspondence Files was changed in 1955 from five-drawer file cabinets to shelf filing equipment. Although this change was prompted by the need to conserve storage space (a ratio of about 2 to 1 in favor of shelf filing space for about half the cost for the same storage capacity when using double stacks), the file clerks observed that shelf filing was less tiring and easier on their nylon stockings. Furthermore, with shelf filing, the clerks can reach a good fraction of the current correspondence without leaving their chairs.

CONCLUSION

There are many designs possible for a correspondence information system. Author filing in combination with a user-oriented retrieval system has many advantages in a research and development environment. These advantages include minimum storage space requirements, correspondence files operable by file clerks without the need for technical supervision, a filing system immediately responsive to over 80% of the requests, and an overall system that processes mail efficiently and quickly.