Searching *Chemical Titles* in the Pittsburgh Time-Sharing System

MARY JANE BLOEMEKE and SIEGFRIED TREU
University of Pittsburgh, Pittsburgh, Pa. 15213

Received February 17, 1969

Current awareness searches of *Chemical Titles*, using the University of Pittsburgh's time-shared computer system, have shown that the nature and the present state-of-the-art of time-sharing render it both unsuitable and inefficient for handling automatic dissemination. A system designed to process relatively short jobs and to offer each on-line user quick-response user system interaction is not simultaneously prepared for the lengthy, continuous processing required for current awareness service.

The Department of Chemistry of the University of Pittsburgh, interested in taking advantage of new information retrieval methods, in February of 1967, decided to begin a current awareness service by automatically searching and disseminating *Chemical Titles*.[1] Every two weeks the subscriber to this service receives a magnetic tape of titles, authors, and bibliographic citations of current articles appearing in some 650 chemical journals. The publisher, Chemical Abstracts Service, also provides programs to search these tapes; the one available in 1967 was for the Basic Programming Support.

The Computer Center of the University of Pittsburgh had, at that time, an IBM/360 Model 50 operating under the Pittsburgh Time-Sharing System (PTSS).[2,3] Since 1967, the Pitt system has progressed until it now includes such processors as the assembler, FORTRAN IV, text editor, and PENELOPE (Pitt Natural Language Processor) and has expanded service to nearly fifty remotely located, time-sharing terminals. Early in 1968, the Operating System (OS) of IBM was made separately available on the same machine to enable standard batch-processing. More recently, the Computer Center has acquired a second 360 Model 50 to provide more extensive and efficient service to batch-processing and time-sharing users simultaneously.

However, in 1967, the first step in beginning *Chemical Titles* service had to be the modification of the CAS program to run under the Pitt Time-Sharing System (PTSS), the only operating system then available. A part-time applications programmer, a first-year graduate student in the Chemistry Department, was found to do this job. Extensive changes were required, and, as a result, the program provided by CAS was rewritten completely. All of the logic and truncation options of the CAS program were retained, with the further capability of allowing nested parentheses so that several logically unrelated questions could be processed for one user profile.[4] The revised program made it impossible to search more than one profile at a pass of the tape. While this was in itself an inefficient approach to searching and undoubtedly contributed to the difficulties here described, it does not negate the comparisons made between time-sharing and batch processing.

EXPERIENCES GAINED

Service to a trial group of 13 users began in August 1967. The time-sharing computer was also new, and, though subject to the usual machine failures, the relatively light use of the system enabled *Chemical Titles* to run quite smoothly. Within a year, however, the number of user profiles increased to around 50. As the peak-period traffic on the system approached the Model 50's saturation point, serious problems developed. These problems were normally attributable to one or more of the following:

1. System hardware or software failure (during CT searching)
2. Excessive console time (search initiation to completion)
3. Console operator (for directing the searches)

Finding a terminal operator capable of learning a series of very simple commands would have been relatively easy. But system failures complicated the job to the point that more than a rudimentary understanding of the typewriter console was needed. To enable the console operator to take appropriate corrective action at time of system failure, some knowledge of the operating system and tape handling procedures was required. It was necessary to monitor the console at all times when search was in progress, and console baby-sitting sessions became increasingly time-consuming and frustrating. By August 1968, the console located in the Chemistry Library was in use about twenty hours per week for 51 profile searches.

From the point of view of a system user, the situation became intolerable. The time-sharing system necessitated the training of a console operator to render him fairly knowledgeable of the system's actions and reactions, as well as the search program. It also required that he have enough interest and time to spend many hours at the terminal. Of even greater concern was the matter of economy. While time of console operation was not figured in time-sharing computer costs, interference and consequent operator error were quite costly, since whatever search processing was already accomplished at the time of involuntary job abortion was normally unrecoverable, both in time and cost. In summary, time-sharing service for CT searching, as seen by the user, was clearly unsatisfactory.

But what about the point of view of the time-sharing system and its staff? In respect to the problem of system failure cited earlier, less than an adequate level of system reliability is generally acknowledged to be troubling most time-sharing installations. Recent national computer conferences have devoted considerable time to discussing this matter. However, until the computer industry overcomes computer hardware and software failures more effectively,

Table I. Summary of Chemical Titles Searches

Issue	CT #4	CT #5	CT #6	CT #7	CT #8	CT #9	CT #10	CT #11	CT #12	CT #13	CT #14
Date	2-20-68	3-20	3-20	3-20	4-19	5-9	6-12	7-30	7-30	8-13	8-14
Time[a]	69.4	135.2	160.7	230.9	224.2	218.3	229.9	260.8	165.2	83.7	83.2
No. of profiles	17	34	34	40	42	43	48	49	38	27	27
No. of terms	412	741	741	892	919	907	1,075	1,046	526	377	392
No. of alerts	660	1,289	1,850	1,789	1,968	1,548	1,365	1,384	1,185	796	797
Appr. no. titles searched	...	...	...	5,170	4,730	4,490	4,216	3,770	4,032	3,886	4,118
Av. time/profile (min.)	4.1	3.9	4.7	5.9	5.3	4.8	4.8	5.3	4.3	3.1	3.1
Av. no. terms/profiles	24.2	21.8	21.8	22.4	21.8	21.1	22	21	14	14	15
Av. alerts/profile	38.8	37.9	54.5	44.7	46.9	35.2	30	28	31	29	29
No. replying	9	19	19	16	18	12	15	20	6	0	0
% replying	53	56	56	40	43	28	33	41	13	...	...
Av. % of relevant titles	40	37	43	50	43	37	42	56	50	...	...

[a] Central Processing Unit time in minutes.

the computer users must continue to cope with them and circumvent them however possible.

The second problem, that of excessive console time, was due only in part to system failures. The Pittsburgh Time-Sharing System was deliberately designed for relatively short work sessions requiring user-system interaction. Since true "time-sharing" implies not only the sharing of central processor time, but also sharing of peripheral devices, e.g., tape and disk units, if numerous users are to be able to work simultaneously on the system, no one should monopolize any system component over extended periods of time. It is thus not surprising that difficulties were experienced by a user attempting long on-line searches, primarily noninteractive in nature, on a system designed for short, interactive work. A temporary system failure is annoying to the person doing conversational work at the console, but is rarely very expensive or difficult to correct. This is not true of a long tape processing job which requires uninterrupted searching until completion.

The nature of time-sharing systems is also such that simultaneous demands by many users can have detrimental effect on actual time required to complete a search.[5] During heavy use of the Pitt system, CT search time increased by as much as a factor of four. These longer search times, in turn, increased the probability of interference due to some system failure.

SELECTED SEARCH STATISTICS

Table I summarizes some of the search statistics accumulated for the period March through August 1968. Included are figures of relevance judgments relayed by feedback cards from March to July by an average of 40% of the profile holders. Average percentage of relevant items per issue of CT was 44%. During the University recess in August, user feedback was not obtained.

The Central Processing Unit time expenditures displayed are not directly convertible to costs. The charges for use of the Pittsburgh Time-Sharing System are based on an algorithm which incorporates three billing factors: central processor (CPU) time, amount of core space employed (during execution), and the number of peripheral devices, e.g., disk or tape units, controlled over a period of time. With this approach, each user is required to pay only for use of his fair share of the computer system, depending on whatever combination of the above three factors he demands. It also means, however, that it is inconvenient to discuss costs, because a fixed "hourly rate" is not applicable. For estimating purposes, however, an average "hourly cost" of $250 can be used.

How does CPU time vary as the number of terms per profile increases? Figure 1, a plot of this relationship, is based on profile searches in issue 10 of *Chemical Titles*. A more detailed description of the terms, logical structure, and truncation modes employed for each profile is given in Table II.

The Research Office of Sociology, one of the participants in the Chemical Information Center at the University of Pittsburgh, performed multivariant analysis upon the factors which could be identified as having a bearing on search times.[6] A high correlation was found between CPU time and the number of terms per profile. Other factors, such as truncation mode and logical structure, were considerably less important.

REMEDIAL ACTIONS TAKEN

Short of a redesign of the PTSS *Chemical Titles* search programs, at least two actions for more successful and economical searching of CT within the Time-Sharing System suggested themselves. An obvious step was the deliberate identification of periods of light system use and the sub-

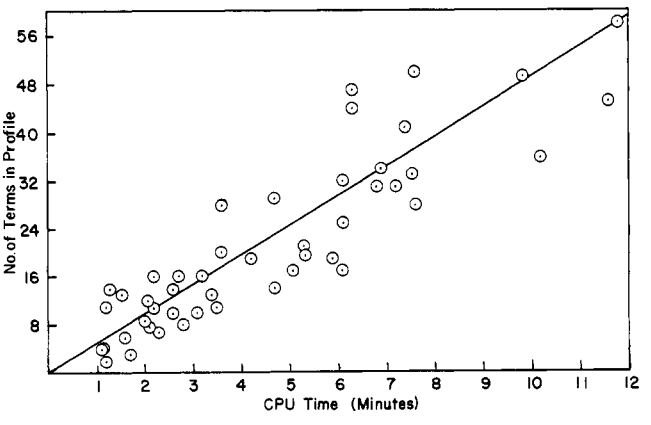

Figure 1. CPU time *vs.* terms in profile for CT # 10.

Table II. Analysis of Profiles Used to Search CT # 10

Terms		
Total number		1,075
Number of truncation	Mode 1[a]	8
	Mode 2	570
	Mode 3	448
	Mode 4	49
Number of not terms		87
Central Processing Unit (CPU) time, minutes		229.9
Profiles		
Number		48
Average number of terms per profile		22.4
CPU time per profile, minutes		4.8
Number of logical structures	A[b]	9
	A and B	13
	A not B	1
	A and B and C	2
	A or (B and C)	16
	A or (B not C)	1
	A or (B and C) or (D not E)	6

[a] Truncation modes 1*---, 2---*, 3---, 4*---*, where an asterisk indicates a blank space or any character. [b] Letters represent parameters of one or more terms connected by *or*.

sequent restriction of CT runs to these periods. This was done to reduce the lengths of console sessions, as well as the chance of experiencing system difficulties. Secondly, based on the relationship between CPU time and the number of search terms, a limit (to 25) of the number of terms per profile was imposed.

However, the remedial actions taken became only interim measures following the August 1968 implementation of the IBM OS system on a separate 360/Model 50 in the Computer Center. It was then decided that current awareness service be shifted from time-sharing to batch-processing as soon as necessary programs could be obtained.

SOME CONCLUSIONS AND EXPECTATIONS

Assuming that the Pittsburgh Time-Sharing System is representative of currently existing time-sharing systems (indications are that it is), then it is safe to assert that the nature and state-of-the-art of time-sharing make it both unsuitable and inefficient for the long tape-searching jobs required by current awareness services. In particular, searching *Chemical Titles* in the time-sharing environment which is prone to system or console operator failure is lengthy and too costly. Furthermore, until time-sharing systems evolve into all-inclusive services effectively catering to the diverse needs of all computer users, it appears that current awareness searching does not belong in a time-sharing system. There is no reason to pay for the luxury of a system which is highly oriented towards user-system interaction, when the repetitive searching based on established user interest profiles is essentially noninteractive.

It is expected that a dual-mode approach, taking advantage of the strengths of both time-sharing and batch-processing, may be most successful in providing present-day chemical information services. Batch-processing, perhaps including remote job entry facility, can continue to be used for efficient processing of the relatively unchanging, regularly scheduled, noninteractive type of work, such as current awareness searching based on standing interest profiles. Time-sharing, on the other hand, holds much potential with respect to satisfying other chemical information needs. Because of the user-system interaction it offers, it can be employed in a specialized "computer-aided instruction" manner for training the users of a chemical information center. An extension of this would be the user's on-line construction and modification of his interest profile.

Although retrospective searching of large files of chemical references is generally lagging (in implementation and usage) behind corresponding current awareness searching, it is reasonable to expect it to become more generally available, and also that time-sharing may play an important role in it. Time-sharing could render valuable assistance in retrospective search (strategy) formulation by greatly reducing time otherwise wasted in processing queries without any indication as to anticipated success and output quantity. Beyond that, the actual on-line retrospective searching of large files for one-time queries is not yet feasible. Considerably more work on file structures and on optimal use of available random access devices will be required. However, since a person with a one-time query (unlike the current awareness person) can be presumed to have motivated interest in obtaining search results as soon as possible, why not compromise and offer "semi-interactive" searching as part of the time-sharing service? This is essentially the approach being developed at Pitt: the user formulates his search request at a time-sharing terminal and stores the finished request for subsequent transmittal (along with other search requests) to the batch-processing computer, at which time efficient, noninteractive searching is carried out.

ACKNOWLEDGMENT

This work was initiated by E. M. Arnett, Professor of Chemistry, and carried out under his direction. The authors are grateful for his interest and assistance throughout.

Financial support for this work has been received from the National Science Foundation under grant GN738 and from the National Institute of Health under Grant No. FR-00250.

LITERATURE CITED

(1) Freeman, R., J. Godfrey, R. Mainzell, R. Rice, and W. Shepard, "Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries," J. CHEM. Doc. **4**, 107 (1964).

(2) "Pitt Time-Sharing System for IBM System/360," Computer Center, University of Pittsburgh, revised March 1968.

(3) Badger, G., E. Johnson, and R. Philips, "The Pitt Time-Sharing System for the IBM System 360: Two Years Experience." Proceedings of the Fall Joint Computer Conference, San Francisco, December 1968.

(4) "Preparation of Search Profiles," Chemical Abstracts Service, Columbus, Ohio, 1967.

(5) Kemeny, J., and T. Kurtz, "Dartmouth Time-Sharing," *Science* **162**, 223 (1968).

(6) Amick, D., "A Multivariant Research Note," Research Office of Sociology, University of Pittsburgh, 1968. (Mimeo)