omitted. References were given to US 3 343 930 and US 3 393 089 which are the two US patents belonging to family 1, the basic invention.

Table VI shows how these other publications were covered by the abstracts journals under study. No reference to GALVALUME was found in *Bull Signal*, CA, or BTI, but only an author search was made through *Bull Signal* and the commercial nature of many of the articles found elsewhere probably accounts for their absence from CA. BTI covers only British technical journals and all the references found were to US or Australian periodicals.

Zn Abs and *Applied Science and Technology Index* included only one of the references, while the two covered by *BNF Abs* were the same article in two different journals. Three relating to the development of ZINCALUME in Australia were monitored by *Zn Bull Anal*, and no one else, in August 1976. *W Al Abs* and *Metals Abstracts* covered the highest number of the references (four each) and were the only two so far to quote the detailed article by Harvey (j). *Metals Abstracts* included it only in June, 1977, so it is possible that other journals will include it in future issues.

The small number of references covered by each journal is probably due in part to the relatively obscure periodicals in which some of the articles were published, such as *Southam's Metalwork* and *Lysaght Brownbuitt Industries*. *W Al Abs* in August 1972 was the first to cite a journal article relating to GALVALUME. By August 1973 all the journals referring to any GALVALUME literature, apart from *Zn Bull Anal*, had mentioned GALVALUME. But this was 6 to 7 years after the appearance of the first relevant patents, and up to 1976, the articles referred to contained few technical details and they did not give the relevant patent numbers for anyone wanting to find out more about GALVALUME.

The assumption made by *Metals Abstracts* that all information contained in patents of importance will be repeated in journals articles has therefore, in this case, been shown to be false. Using journals only, the existence of GALVALUME would have been realized 6 years after the first published patents were available and, without referring to the patents, details of the process by which it is formed and used would still not be known. These results also confirm those found by Liebesny[3] that items appearing in the patent literature rarely also appear in the nonpatent literature.

CONCLUSIONS

That patents are valuable and often unique sources of information tends to be stated or rejected with little evidence to back it up. Thus, for example, when Liebesny stated[4] that patents are frequently the only source of information, give more information than any other form of literature, and frequently give information much sooner than any other form of literature, he based this on his own subjective experiences rather than on hard evidence. We believe our results confirm Liebesny's views in one particular subject area, and we hope they have gone some way to convincing people who think that patents are not worthwhile sources that they need to reconsider their views.

ACKNOWLEDGMENT

We would like to thank Mr. Martin White and Ms. Bridget Hayes of Zinc/Lead Development Association for their assistance.

REFERENCES AND NOTES

(1) C. Oppenheim and E. A. Sutherland, "Studies on the Metallurgical Patent Literature. I. The Coverage of Patents by Abstracts Journals in Metallurgy", *J. Chem. Inf. Comput. Sci.*, preceding paper in this issue.
(2) T. Graff, personal communication.
(3) F. Liebesny, J. W. Hewitt, P. S. Hunter, and M. Hannah, "The Scientific and Technical Information Contained in Patent Specifications", *Inf. Sci. Abstr.*, **8**, 165–177 (1974).
(4) F. Liebesny, "Mainly on Patents", Butterworth, London, 1972. Chapter 6, pp 120–121.

"The Merck Index": The Merits of Using Computers in Publishing†

MARTHA WINDHOLZ* and HORACE D. BROWN

Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065

THOMAS G. GASPAR

Merck & Co., Inc., Management Information Systems, Rahway, New Jersey 07065

Received November 1, 1977

"The Merck Index", an internationally recognized encyclopedia of drugs, chemicals, and biologicals was produced by the traditional method for eight consecutive editions. Owing to the rapidly increasing volume of information and escalating costs of book production, it became desirable to design a computer-assisted production method for preparation of the recently published ninth edition. The method that was developed and used as well as the philosophy behind its use are described.

"The Merck Index", an encyclopedia of chemicals, drugs, and biologicals has been published periodically by Merck & Co. since 1889. In its original form Merck's Index was a 170-page alphabetical list of E. Merck products sold all over the world. Through an evolutionary process and several successive editions, "The Merck Index" has developed from a company catalog into an authoritative, internationally recognized, 2000-page reference source with a circulation of about 150 000 copies per edition. It contains capsule descriptions (monographs) of about 10 000 chemicals, drugs, pesticides, and biologically active compounds, together with

† Presented in part before the 6th Annual Meeting of the American Society of Indexers, New York, N.Y., April 29, 1977.

0095-2338/78/1618-0129$01.00/0

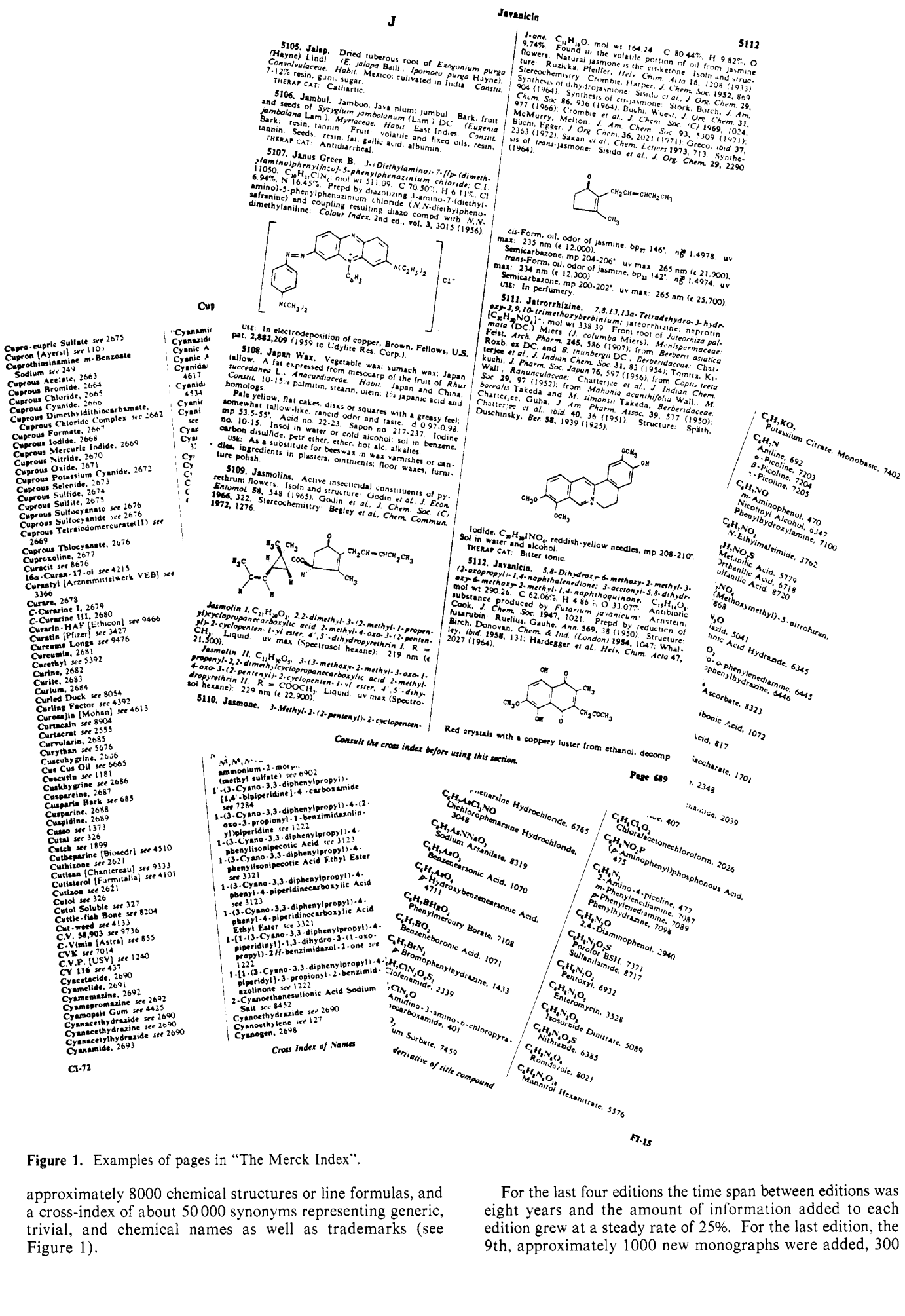

Figure 1. Examples of pages in "The Merck Index".

approximately 8000 chemical structures or line formulas, and a cross-index of about 50 000 synonyms representing generic, trivial, and chemical names as well as trademarks (see Figure 1).

For the last four editions the time span between editions was eight years and the amount of information added to each edition grew at a steady rate of 25%. For the last edition, the 9th, approximately 1000 new monographs were added, 300

monographs were deleted, and about 50% of the text was revised and updated.

Through the 8th edition, published in 1968, "The Merck Index" was always produced by the traditional manual methods.

TRADITIONAL METHOD

The process of copy (manuscript) preparation is difficult to describe because practically each monograph (and there are 10 000 of them) is handled differently. New information is assembled continuously by scanning and correlating data from literally hundreds of sources from all over the world. Monographs are written by a staff of subject specialists which, for the 9th edition, consisted of four chemists and three technical assistants. It is not unusual for at least half of the monographs to go through three to four different revisions, and as many retypings, before reaching their final form. All typing is done by typists engaged solely in the preparation of new and clean manuscript which is carefully proofread and copy edited for the typesetter who then retypes the text for printing using the monotype method. At this point errors are unavoidably introduced, and new effort on the part of the editorial staff has to be exerted (in the form of at least three proofreading cycles) to detect and correct all errors before the text is ready for printing.

The whole process is probably best illustrated by an analogy with the preparation of a scientific article for publication; but while the average length of an article is 20–30 manuscript pages, in this case the editors are dealing with a mammoth 6000–7000 page manuscript.

In addition to manuscript preparation, there is the most time-consuming task: index preparation. For any single edition, approximately 50 000 names and empirical formulas embedded in the text are extracted manually, carefully alphabetized, typed, and proofread; they are then retyped by the typesetter and reproofread by the editorial staff before page numbers can be assigned to them (see Figure 2).

It was during the production of the 8th edition, in the late '60's, that it became evident that the traditional method had outlived its usefulness. The effort of index preparation alone took about 24 man months. Transition from original manuscript to final page and bound book had become a 17 calendar-month process. With the rapid progress in science and the exponential growth of scientific publications, the editors were convinced that other alternatives to the traditional method would have to be investigated.

SEARCH FOR ALTERNATE METHODS

Computer-assisted composing and printing methods emerged in the early '60's. As the 9th edition was being planned, the editorial staff was studying the significant technological advances in the printing industry which afforded the capability to store, maintain, and edit manuscript captured in machine-readable form for subsequent printing.[1,2] The new computer-assisted editing systems were known to be best applicable to databases rather than to texts printed on a one-time basis.[3,4]

Could "The Merck Index" be thought of as an ever-changing database rather than a mammoth scientific manuscript? Encyclopedic in nature, with approximately 40% of the information contained in it being repeated from edition to edition, with the process of production consisting of adding new entries, updating or deleting old entries, "The Merck Index" closely matched the description of a database.

This fact, coupled with the experience that production of an encyclopedia consisted of two distinct phases: (1) an intellectually stimulating, highly creative phase—manuscript preparation—and (2) a painstakingly tedious, time-consuming, and expensive phase—transition from manuscript to final page—made the possibility of computerization extremely attractive and desirable.

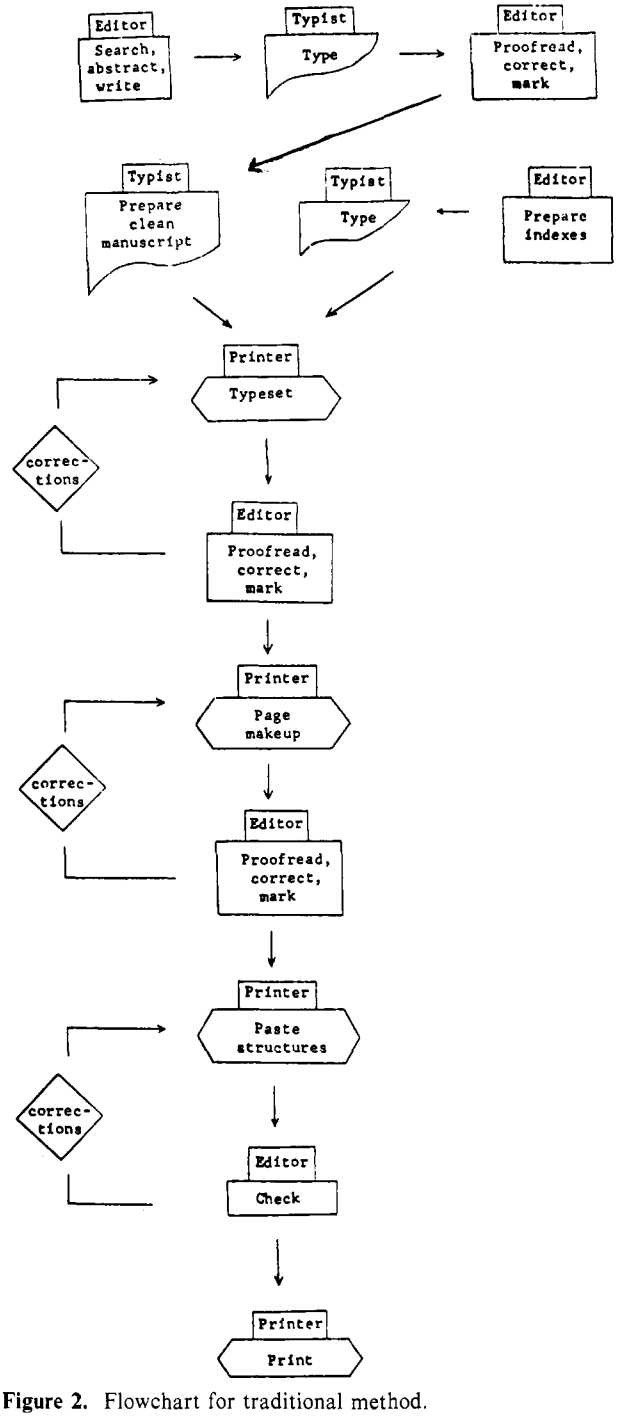

Figure 2. Flowchart for traditional method.

The ideal system imagined was one that would separate the intellectual processes from the purely mechanical ones, or in other words, one that would subordinate the high-speed machine processing of information to the intellectual effort involved in the compilation of an encyclopedia. It would simplify the least creative processes of editorial work, streamline the editing and composing of the book, and, at the same time, achieve shorter schedules and possibly lower costs throughout the publishing process. Thus it would solve two major problems: (1) free the editors from the tedium inherent in book production and (2) reduce the time elapsed from finished manuscript to final page.

A reconnaissance study conducted by the Automation & Control Department at Merck & Co., Inc., established the

availability and economic feasibility of systems, equipment, and services offered in computer-assisted composing and evaluated their applicability to a text as complex as "The Merck Index".

The findings of the reconnaissance study were as follows:

(1) Computer-assisted techniques and equipment for text editing and composition were being used routinely in the industry not only for fast turnover and large volume printing operations such as newspapers,[5,6] journals,[7] and directories, but also for more esoteric publications such as *Chemical Abstracts*,[8,9] Blakiston's "New Gould Medical Dictionary", "American Heritage Dictionary", etc.

(2) The existing techniques were applicable to the composition or typesetting operations related to "The Merck Index" database.

(3) An automated system would reduce by 25% the cost of preparing an edition of "The Merck Index" for the printer, more than offsetting the cost of computer development. The savings would come from substituting an automatic process for conventional manuscript typing, indexing, type specification, and composition. Moreover, repetitive proofreading would be eliminated and the number of typographical errors would be greatly reduced.

(4) Automation would allow more frequent publication of a more up-to-date encyclopedia.

(5) Automated production of subsequent editions of "The Merck Index" would result in further savings since about 40% of the text would not have to be retypeset and reproofread.

In order to design a system and implement the computer-assisted method which was used to produce the 9th edition, a multidisciplinary task force consisting of chemist-editors, systems engineers, and programmers was established in 1972.

SYSTEM AND COMPUTER-ASSISTED METHOD

The system with which the ninth edition was produced includes a text-editing and text-setting terminal (Imlac PDS-1), a central information storage and retrieval computer (CDC-6400), and a photocomposer (RCA Videocomp).

The editing terminal consists of a computer and two television-like video display tubes, one equipped with a keyboard for entering the manuscript (ca. 10000 monographs) into the system and the other equipped with a camera for producing printed copy text.

The manuscript is compiled the same way for the computer-assisted method as for the traditional method. To enter the manuscript, the operator responds via the keyboard to a questionnaire displayed on the video screen. The questionnaire elicits specified elements of information contained in the monographs such as title, alternate name, empirical formula, molecular weight, literature references, etc. (Individual monographs, representing one record each, are entered in a random order.) When all questions pertaining to a monograph have been answered, the equivalent of a page-proof can be displayed on the video screen by pressing a button. The system has specified the type and done the composition. Structuring the text according to a routine permits the system to label standard elements of the text automatically. This, in turn, provides for automatic preparation of the index and will provide, at a later date, instantaneous retrieval of unique items.

All characters typed by the operator appear on the screen exactly as they will look in print, whether they are upper case or lower case letters, subscripts or superscripts, bold face or light face, etc. The system provides for entry and display of over 400 symbols, including Roman, Italic, and Greek characters, and chemical and mathematical symbols, without the operator having to type codes for bold face, upper case, new paragraphs, or to strike a key for carriage return.

When the composed monograph is on the video screen, the operator checks it for errors, proper hyphenation, and right-hand justification. The terminal is then instructed to produce a printed copy for the use of the editorial staff after which an electronic copy of the monograph is transmitted to the central computer for safekeeping.

The editors proofread and correct the printed copies and return them to the operator who recalls the text from the central computer and enters the corrections at the editing terminal. The computer incorporates the corrections, adjusting the composition according to the programmed specifications for spacing. The text can then be redisplayed and returned to the central computer for storage.

The central computer stores and safeguards the "proof" of each monograph. It also is programmed to sort monograph titles alphabetically, to number them sequentially, and to compile an index of all synonyms and empirical formulas entered into the database.

The indexing process involves the examination of about 50000 terms (tagged by the operator on the editing terminal during keyboarding), alphabetization of these terms according to special rules, and their coupling with the pertinent monograph numbers. The completed index can be displayed on the editing terminal's video screen for final review.

When editing and correction of all three sections of the book—monographs, formula index, and cross reference index—are completed, the copy and the composition instructions contained in the central computer are transferred to a series of magnetic tapes which are turned over to the composition supplier[10] for photocomposition.

The magnetic tapes are translated to typesetting standards incorporating all information needed for page formatting: vertical and horizontal line justification, illustration spacing, running heads, running feet, telltales, and page numbering. The new tape obtained drives the *photocomposer* (RCA Videocomp) which produces camera-ready copy. All illustrations are then manually stripped into the spaces provided by the computer. The finished pages are checked for accuracy by the editorial staff and released to the printer.

The entire original manuscript is retained by the central computer and will be used as a basis for the next edition (see Figure 3).

Comparison of the flowcharts in Figures 2 and 3 clearly demonstrates that the use of computers provides the editors with a convenient tool to do extensive manuscript revision and composition without having to go through the endless editing cycles of rechecking.

Today "The Merck Index" composition system is a fully operational production system. Its design, development, and implementation lasted about three years. The advantages of combining the speed and mass storage capability of computers with the most advanced photocomposing equipment were analyzed and are summarized in Tables I and II.

Table I. Manpower Usage (in man years) for Traditional and Computer-Assisted Methods[a]

	Traditional system	Computer-assisted system
1. Manuscript preparation: research writing, typing/typesetting	26	26
2. Proofreading galleys	8	
3. Proofreading corrected galleys	1	
4. Proofreading pages	0.16	0.16
5. Preparation of indexes	5	
Effort from galleys to final page (nos. 2–5)	14.16	0.16
Total effort, from manuscript preparation to final page	40.16	26.16

[a] Exclusive of manpower used by Merck for system development and by composer for typesetting by monotype method.

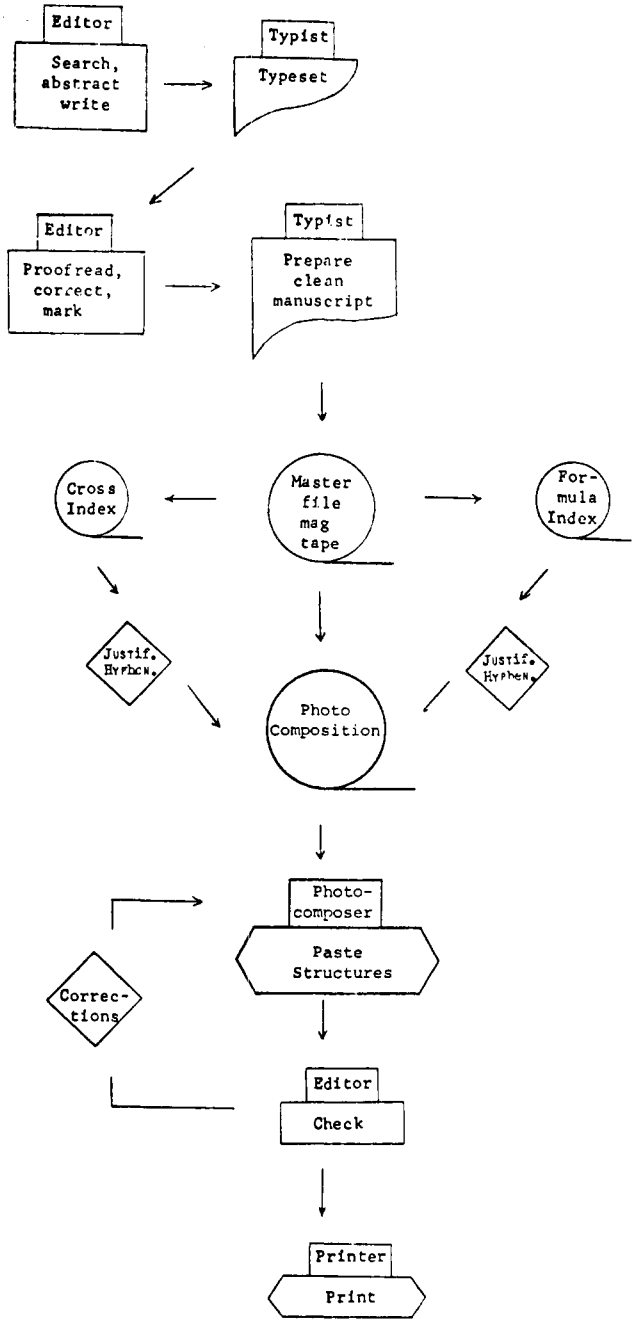

Figure 3. Flowchart for computer-assisted method.

Elimination of costly and time-consuming index preparation, monotype typesetting, and unnecessary proofreading cycles resulted in net savings of 14.0 man years (~35%) of editorial effort and reduced the time elapsed between freezing the manuscript and actual publication date from 17 months to 4 months.

The results achieved are quite significant in view of the facts that: (a) full versatility of the system could not be exploited the first time around, because the entire text of the encyclopedia had to be keyboarded and proofread; and (b) full economic advantages could not be realized because the expected reduction in typesetting cost was offset by computer development cost.

Table II. Elapsed Time (in years) from Finished Manuscript to Bound Book for Traditional and Computer-Assisted Methods

	Traditional system	Computer-assisted method
Typesetting, galley correction	1	
Video composition		0.04
Page makeup and structure pasting	0.17	0.04
Printing and binding	0.25	0.25
Total time elapsed from finished manuscript to bound book	1.42 (17 months)	0.33 (4 months)

A larger decrease in production costs is anticipated on future editions when only additions, deletions, and changes to the master file will have to be entered to bring the encyclopedia up to date.

The computer-assisted method also offers the following advantages: changes in the text can be made on a daily basis, as new information becomes available; only updated portions of the book need to be proofread; and changes can be incorporated up to the last minute. Moreover, with minimal further system development, the possibility of having a fully updated encyclopedia ready for printing, practically every week of the year, is soon to become reality.

ACKNOWLEDGMENT

One of the unique features of "The Merck Index" is that it is not edited and published by a publishing house but by Merck & Co., Inc. Research and development of methods leading to the use of computers in publishing normally do not belong in the domain of a company specializing in drug research. Therefore, we owe a big debt to the management of Merck & Co., who, in spite of the fact that "The Merck Index" is a nonprofit venture, was willing to support this pioneering effort not only with funds but also by assigning to this project a team of highly competent computer experts.

The outstanding contribution of "The Merck Index" editorial staff and of the Automation & Control Department staff is gratefully acknowledged.

REFERENCES AND NOTES

(1) M. P. Barnett, "Computer Typesetting, Experiments and Prospects", The MIT Press, Cambridge, Mass., 1965.
(2) "Editing/Typesetting Systems: a Review of a Fast Moving Art", *Publisher's Weekly*, 39 (July 3, 1972).
(3) "Producing a $4 Million Book", *Publisher's Weekly*, 1 (Sept. 1, 1969).
(4) "Computers Make Steady Inroads in Book Editing and Composition", *Publisher's Weekly* 35 (Dec. 4, 1972).
(5) N. Mitchell, "Proofing Copy in the Computerized Newspaper", *Modern Data*, 22 (Aug 1974).
(6) J. Walsh, "New York Times: All the News That's Fit to Printout", *Science*, **181**, 640 (1973).
(7) F. L. Alt and J. Y. Kirk, "Computer Photocomposition of Technical Text", *Commun. ACM*, **16**(16), 386 (1973).
(8) "ACS Journals Plan Switch to Computerized Photocomposition", *Chem. Eng. News*, 24 (Nov 20, 1972).
(9) N. A. Farmer, "The American Chemical Society Composition System", CAS Report No. 5, Oct 1975, p 3.
(10) Photocomposition of "The Merck Index", 9th edition, was carried out by International Computaprint Corp., Fort Washington, Pa.