DOCUMENT RESUME

ED 092 114 IR 000 673

AUTHOR Gill, Elizabeth Deas
TITLE A Comparison of Manual and Computer Searches of the Chemical Evolution and Origin of Life Literature.
PUB DATE Jan 74
NOTE 53p.; Master's Thesis, San Jose State University

EDRS PRICE MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS Bibliographies; Biology; Chemistry; *Comparative Analysis; *Computers; *Evaluation; Evolution; *Information Retrieval; *Search Strategies
IDENTIFIERS *Manual Searches

ABSTRACT

Parallel machine and manual literature searches on the subject of chemical evolution and the origin of life were compared on six characteristics: (1) precision, (2) recall, (3) novelty, (4) uniqueness, (5) time cost per citation, and (6) dollar cost per citation. The manual search outperformed the machine on precision, novelty, uniqueness, and dollar cost per citation although this was based on partial cost data for the manual search. There was little difference in recall between the two methods. For this subject area, "Chemical Abstracts", "International Aerospace Abstracts", and the Automatic Subject Citation Alert service were found to be the most effective sources for overall recall precision, novelty, and uniqueness. RECON and MEDLARS were found to be the most efficient in terms of times and costs. The study concludes that while the manual search had a slightly better overall performance, both modes are necessary for a comprehensive multi-disciplinary literature survey. (JG)

A COMPARISON OF MANUAL AND COMPUTER SEARCHES OF THE CHEMICAL EVOLUTION AND ORIGIN OF LIFE LITERATURE

by

Elizabeth Deas Gill

January 1974

A COMPARISON OF MANUAL AND COMPUTER SEARCHES OF THE CHEMICAL EVOLUTION AND ORIGIN OF LIFE LITERATURE

A Research Paper

Presented to

The Faculty of the Department of Librarianship

California State University, San Jose

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Elizabeth Deas Gill

January 1974

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

OBJECTIVE:

The purpose of this study was to compare the results of parallel machine and manual literature searches in the preparation of an ongoing bibliography. This bibliography is comprehensive and multidisciplinary on the subject of Chemical Evolution and the Origin of Life. The specific objectives of this work were to determine:

1. which source(s) is the most effective using the User criteria of recall, precision, novelty and uniqueness
2. which source(s) is the most efficient using the Management criteria of time and cost
3. possible ways to economize or improve future searches.

VALUE OF THE STUDY:

In spite of the increase in automated information retrieval systems, little study has been done to determine how effectively these systems can replace or aid the traditional methods of search. The vast majority of the studies done have been on the relative value of indexing procedures within one system. When systems have been compared, the objective of the study has been a very small synthetic search of a current awareness nature, focusing on one or two characteristics only. In these days of tight money and rising costs, management needs quantitative data about how best to use the tools and services available to them in real-life situations. It is especially important with ongoing projects, such as the bibliography under study, to know what is being done and what can be done more effectively and or efficiently. An analysis of multiple characteristics of a real and ongoing, retrospective literature search has not been undertaken to date. It is felt that the Chemical Evolution and Origin of Life Bibliography provides an acceptable basis for such a study.

Each year for the past three years, the Chemical Evolution Branch, Life Sciences, NASA-Ames Research Center has commissioned the updating of the basic bibliography, "Chemical Evolution and the Origin of Life," compiled by Martha W. West and Dr. Cyril Ponnamperuma and published in Space Life Sciences 2(1970) 225-295. These supplements include all literature in the subject field published the previous year and found in several scientific disciplines - astronomy, biology, chemistry, geology, etc. The circumstances of this literature search seemed ideal in size, scope, and subject array for a comparison of the various criteria mentioned earlier. In addition, the conditions of the literature search are duplicated in many special libraries all over the country -

1. a library well-stocked with the indexes and abstracting services geared to the mission of the parent agency
2. availability of several commercial computerized search services
3. an in-house information retrieval system
4. a literature Searcher with both library and subject background.

With these aggregate conditions present, it is hoped that the results of the current study can fill a void in the evaluation literature. Equally important, it is hoped that this study will provide an objective basis for User and Management decisions in economizing future bibliographies as well as aiding the Searcher in more proficient surveillance of the literature.

REVIEW OF RELATED WORK:

Studies of many characteristics of retrieval systems (computerized and manual) have been made but efforts have concentrated on the input or indexing stage. When attention has been focused on the search or output stage, the work has been mostly an evaluation of indexing systems within the system itself - rarely as a comparison with another system, manual or computerized. For example, Cleverdon,[1] Lancaster,[2] and Salton,[3] in their evaluations, arrive at complicated ratios of recall, precision, etc. for various indexing methods within a single information retrieval system but never do they answer the question, "does this system as a whole retrieve information more completely, faster or less expensively than that system?"

Within the manual search mode, Spencer[4] compares the preparation of a drug bibliography using Chemical Abstracts and Index Medicus for the years 1956-1964 with the results of another using Science Citation Index 1961 and 1964. An equal length of search time was used with each tool and the results were compared in various areas. She found that each of the three indexes produced a high percentage of unique references but no appreciable differences in efficiency. But when the total search time was limited to very short time spans, Science Citation Index performed much more profitably than did Chemical Abstracts or Index Medicus. She made no attempt to generalize her conclusions.

Within the computerized mode, emphasis has been on current awareness services rather than retrospective searches. In 1971, Barker, Veal and Wyatt[5] compared the relative performance of four major Chemical Abstracts Service magnetic tape data-bases in terms of relative currency and retrieval efficiency. Data-bases varied from "titles only" to "titles plus keyword phrases" to "titles plus abstracts." Fifty questions were used and profiles for each were matched against each of the data-bases. Precision and recall percentages were calculated for each question. Results were as they expected - the recall increased as the volume of searchable material per citation increased.

When the comparison is between manual and computerized searches, the emphasis, again, has been on current awareness tools or MEDLARS. H. P. Angstadt[6] of the Sun Oil Company compared the output of computer searches of Chemical Titles and Chemical Abstracts Condensates against the product of several searches scanning the same material by their normal procedures. Twenty-eight single question profiles were prepared and used as the basis of the study. Only computerized results were evaluated in terms of recall and precision. A second set of results was calculated after a revision of most of the profiles. Once optimum profiles were developed, Angstadt found no appreciable differences in precision and recall values for the two computerized searches.

Chemical Titles was, also, the subject for a study done by Lynch and Smith[7] at the University of Oxford in 1970. This investigation compared the Chemical Titles output on one carefully constructed profile over a period of a year with a manual search of the relevant literature. Results were reported in numbers rather than percentages. Areas of comparisons were very limited - only those items found by both computer and manual searching received any degree of statistical manipulation. Relative speed of retrieval was the important factor in this study.

Another study of automated current awareness services versus manual selection was done in 1970 by R. H. Searle[8] of the United Kingdom Atomic Energy Agency. He compared references retrieved from Chemical Titles and ASCA with material retrieved by information and scientific personnel using the primary literature. He found that, at present, the computer-based services using small, high relevance profiles could retrieve only about 50% of those retrieved by human selection.

The most thorough comparison studies have been done with the printed Index Medicus and the computer-based MEDLARS. In 1967, Ohta[9] compared the results of seven medical bibliographies prepared with the use of both tools. Her study indicated that, at present, neither method gave adequate output and results varied so greatly from bibliography to bibliography that no general conclusions could be reached.

MEDLARS and Index Medicus were again the basis of a portion of a study done in 1968 on the literature of Ophthalmology. Both Virgo[10] and Miller[11] reported and interpreted data generated by the study. Different methods of calculating precision percentages led to dramatic differences in results. Still, both reached the same general conclusion that MEDLARS searches are not clearly superior to manual searches of Index Medicus and that for extensive retrospective searches more than one secondary source must be used.

DEFINITIONS OF TERMS AND ABBREVIATIONS

Data-base... ....Information stored by an information retrieval system, in printed or machine-readable form.

Demand search...............Search conducted in response to a specific demand.

Novelty.....................The ability of a retrieval system to reveal relevant information to the User for the first time.

Precision...................The ability of a retrieval system to reveal only information of interest to the User and to hold back unwanted information.

Recall......................The ability of the retrieval system to uncover all relevant information stored in the data-base.

Reference...................The bibliographic data retrieved by searching the eight data-bases of this study, used interchangeably with "item" and "citation".

Relevant item...............Item, citation or reference useful to the User in relation to his information need which prompted his request.

Search......................Action taken from the receipt of an information request to the retrieval of the final document or its citation.

Search Component............One of the eight information systems searched as part of the basic literature survey of this study.

Search Strategy.............The translation of a request into the language of the information system used.

Uniqueness..................The ability of a retrieval system to reveal relevant information not revealed by another retrieval system.

User........................The requester of information.

Profile.....................The words or phrases which describe the interests of the user.

Noise.......................Unwanted or nonrelevant information retrieved by an information retrieval system.

METHODOLOGY:

The basis of this study was a ready-made situation. The annual bibliography updating "commissioned" by NASA-Ames was chosen as the search question since it involved rather broad subject areas and had proven user-value; therefore, it filled the criteria of a real search situation, not a synthetic one. The study was confined to those references bearing a 1972 publication date. The experiment was extended through the first six months of 1973 to allow for time lag of data input. Calendar time span was not a factor of concern in this study.

A conference between the User, the Searcher and a Relevancy Judge was held to set parameters for the search. A comprehensive list of relevant indexing terms was compiled by the Searcher (Appendix A). This list was then manipulated to fit the vocabulary of each data-base searched. Eight services were chosen as information systems most likely to yield relevant references. The four printed sources selected for manual searching provided abstracts as well as titles of indexed literature. The four computer-based services gave "titles only" and were selected essentially because they were the ones available or subscribed to by the User. The eight information systems used in the preparation of the bibliography are listed below. Special characteristics of each as related to the conduct of the study are described briefly.

Computerized Sources:

1. ASCA (Automatic Subject Citation Alert). This is a weekly commercial current awareness service published by Institute for Scientific Information. A personal profile is used to search a machine readable data-base containing natural science research literature from

over 2500 source publications. Entry to the data-base may be made by author, subject or corporation. The profile submitted by the User (Appendix B) was accepted by the Searcher, although a few changes were suggested to produce a higher precision ratio in the future. Because this was a retrospective search, the changes affected only a small portion of the ASCA output used for this study.

2. BIOSIS Standard Profile #812 (Biochemical Evolution). This is a commercial current awareness service published monthly by Biosciences Information Service. A standard profile is used to search a machine readable data-base containing bioscientific research literature collected from approximately 8000 source publications. This data-base is the complete index files of each issue of Biological Abstracts and Bioresearch Index. Entry is by author, subject, biosystem or a Cross index. Profile #812 in the broad field of biochemical evolution is predetermined by the publishers of this service and is designed to interest more than one scientist. This Searcher, therefore, had no control over the search terms used.

3. MEDLARS (Medical Literature Analysis and Retrieval System). This is an off-line search service of the National Library of Medicine (NLM). A profile (Appendix C) designed by the Searcher is used to search a machine-readable data-base containing medical literature collected from over 2400 journals.

Entry is by subject and author. Periodic updating was made by MEDLINE to allow for a six month time lag of data input. MEDLINE is an on-line search service of the NLM. Its data-base is limited to approximately 1000 of the most heavily used journals indexed for MEDLARS for the last three years.

4. RECON (Remote Console). This is a real-time, on-line, time-shared computerized information retrieval system operated by NASA. The data-base contains aerospace research literature collected from government, industrial and academic researchers. STAR and IAA together serve as announcement bulletins for data input into this system. Entry is by author, subject, and corporate author. The profile used for this study was designed by the Searcher (Appendix D).

Manual Sources:

1. BA (Biological Abstracts). This printed abstracting and indexing service is published twice monthly by Biosciences Information Service. It indexes biological, medical and biochemical literature from almost 8000 source publications. Each issue has author, permuterm subject, biosystematic and Cross indexes. Each volume (two per year) has a cumulative index. The Searcher used cumulative indexes for those volumes available. Otherwise only the subject index, Biological Abstracts Subjects in Context (B.A.S.I.C.), of the individual issues was used. Bioresearch Index, a companion service with BA was not one of the information systems included in the survey, but is considered in

certain conclusions reached later. It is published monthly by Biosciences Information Service and provides approximately 100,000 additional citations (not abstracts) annually of items not found in BA. These include symposia, review journals and books, government reports, etc. The indexing system is the same as that for BA. The complete index files of BA and Bioresearch Index are the data-base searched by the BIOSIS Standard Profile Service.

2. CA (Chemical Abstracts). This is a printed abstracting and indexing service published weekly by Chemical Abstracts Services and covers chemical and chemical engineering research literature from approximately 12,000 source publications. Each issue has a keyword subject, author, and patent index. Cumulative indexes are issued for each volume (two volumes each year). Cumulative subject indexes were never available at the right time so all searching was done in the keyword indexes of individual issues.

3. IAA (International Aerospace Abstracts). This printed abstracting and indexing service is published twice each month by the American Institute of Aeronautics and Astronautics, Inc. It covers published literature in the field of aeronautics and space science and technology. Each issue has five indexes - classified subject, personal author, contract number, report number and accession

number. Cumulative indexes are published semiannually and annually. The cumulative subject indexes were used when available. Otherwise, the subject indexes of individual issues were used. The RECON data-base includes all items announced in this publication.

4. STAR (Scientific and Technical Aerospace Reports). This is a printed abstracting and indexing service issued semimonthly by NASA. It covers the report literature on the science and technology of space and aeronautics. Each issue has five indexes - classified subject, personal author, corporate author, contract number, and accession number. Cumulative indexes are published semiannually and annually. Semiannual subject indexes were used for one volume only. Subject indexes for individual issues were used for the other two volumes searched. The RECON data-base includes all items announced in this publication.

All of the sources were international in scope. Whenever possible, the output of computerized sources was handled first in order to prevent bias from having seen abstracts in the manual sources.

The machine literature search started at the end of February, 1973 and was concluded in July, 1973. The complete 1972 output for ASCA and the BIOSIS Standard Profile service were immediately available and a MEDLARS demand search was requested immediately along with monthly updates. The RECON search was delayed as long as possible so the necessity for several updates with high duplication of references would be minimized. Manual searching was begun during the first week of April, 1973. The entire 1972 output of each information system (manual or computerized) was searched at one time. Output after January 1, 1973 was searched for 1972 references as it arrived at NASA-Ames Life Sciences Library. Searching was continued through June 30, 1973 output to allow a six-month time lag for data input. As noted previously, the end product of this study was an exhaustive bibliography so the amount of time devoted to each source was not limited.

When the gross searching was completed, the User's literature file was checked for duplication of references. This data was necessary for calculating novelty and recall percentages.

When manually searching printed sources, only subject indexes were used. Since time was not limited and degrees of relevancy were not assigned, all subject terms were considered equal and were searched in alphabetical order. When entries under relevant terms were found, abstract numbers were jotted down and consulted after all index searching was completed. Only those items thought to be relevant or possibly relevant were then "retrieved". Very few nonrelevant abstracts were consulted in STAR, IAA and BA since full or partial titles are given in their indexes.

As items were retrieved either from computer printout or printed index, bibliographic data was transferred to 4" x 6" cards using the format of the final bibliography. The source information was encoded on the rear of the card. For example, an item found in the ASCA printout was marked "ASCA 6 Mar 73 p4" or "CA 73:92255y" if found in CA. These cards were alphabetized daily and as duplication arose, source information was cumulated on the rear of a master card. Duplicate cards were then removed to another file for other uses.

A Search Record was kept for each source giving volume and issue numbers searched, subject headings used, and time expended. In the case of computer printouts with no volume or other identifying numbers, printout dates were used. The time-keeping procedure was altered as the study progressed (see Chapter 2, Unit Time Characteristic).

All references thought to be relevant or possibly relevant by virtue of title or abstract were retrieved. Full texts of as many items as possible were retrieved after the search was completed. The holdings of the NASA Libraries, San Jose State University Library and Stanford University Libraries were used. In some instances, interlibrary loans were obtained from other libraries. If the item was in a foreign language, the Searcher used English translations or English abstracts by the author, if available. In some cases, the Searcher was able to translate enough to determine relevancy. If there was still a question of relevancy, the Relevancy Judge made the final decision. All items judged not relevant were placed in a nonrelevant file so that the data could be available for this study.

CHAPTER 2

OBSERVATIONS AND RESULTS

General:

The literature search produced 2701 references from all sources, 2677 from the eight sources treated in this study. Of the retrieved items, 280 were considered relevant and included in the bibliography update; 230 of the 280 were retrieved by the eight sources. As shown in Table I, 26 references considered relevant by the User were not retrieved by the search. These were supplied by the User's personal file. All but 7 were articles in three newly published books and most probably victims of time lag. (Contents of two of the books were retrieved by the July and August BIOSIS Standard Profiles.) Hopefully, these 19 items will be retrieved in the next annual update. The 7 remaining were in the geological field. Failure to retrieve them could have been due to inadequate capabilities of the Searcher in that subject area or lack of coverage by any of the databases searched.

Table I. Analysis of 1972 Published References

Total References Retrieved by all sources = 2727

Retrieved from 8 sources of study = 2677	Retrieved from other sources* = 24	Supplied by user = 26
Total relevant = 230	Total relevant = 24	Total relevant = 26
Unknown to user = 154	Unknown to user = 20	

*references retrieved from current journals; bibliographies of relevant articles; publisher's catalogs, etc.

For the purposes of this study, the assumption was made that all relevant references found by the Searcher plus the 26 items in the User's file but not found by the Searcher constituted the entire segment of 1972 literature in the field of Chemical Evolution and the Origin of Life (280 references). Formulas advanced by F. W. Lancaster[2,12] of the University of Illinois Graduate School of Library Science were used to calculate percentages of recall, precision, novelty, uniqueness and time cost.

RECALL AND PRECISION CHARACTERISTICS:

The recall performance of a system is expressed quantitatively as the ratio of the number of relevant references retrieved by that system to the total number of relevant references in the literature. The precision performance of a system is expressed quantitatively as the ratio of the number of relevant references retrieved by that system to the total number of references, relevant and nonrelevant, retrieved by that system. Table II tabulates recall and precision values for each of the eight sources (or search components), which are grouped according to the search mode - manual or computerized. For example:

$$\text{ASCA Precision } \% = \frac{\text{number of relevant references retrieved}}{\text{number of references retrieved}} \times 100 = \frac{109}{1518} \times 100 = 7\%$$

$$\text{ASCA Recall } \% = \frac{\text{number of relevant references retrieved}}{\text{total number of relevant references in literature}} \times 100 = \frac{109}{280} \times 100 = 39\%$$

Table II. Recall and Precision Characteristics for Search Components

Component or Source	Retrieved	Relevant	Precision	Recall
ASCA	1518	109	7%	39%
RECON	210	84	40%	30%
BIOSIS Standard Profile	426	31	7%	11%
MEDLARS	238	30	13%	10%
BA	24	20	83%	7%
CA	137	110	80%	39%
IAA	121	117	97%	42%
STAR	3	3	100%	1%

Column 2 of Table II shows a range in the number of retrieved references from ASCA's high of 1518 to STAR's low of 3. In past years STAR has yielded decidely more citations. Since the U. S. Government has curtailed many of its research programs, the report literature indexed by STAR has decreased in several areas. The increased number of items retrieved by IAA reflects the switch to journal literature.

Only ASCA and BIOSIS Standard Profile had recall values which were greater than their precision percentages. The two services produced over 50% of the total references retrieved by the eight sources. This has resulted in very low precision percentages of 7% each. ASCA's extraordinarily large yield of references is due in part to the sophistication of the service. This tool can search for relevant literature by specific word, phrase or stem, by author, corporation or source, or by cited work or author. The latter means the system can search for any item which cites a particular reference or the work of a particular author. But a badly designed profile can result in the retrieval of a lot of "noise." For example, the profile for this search (Appendix B) has directed the computer to retrieve work by Miller as well as items by other authors who cite Miller. It also directs the computer to do the same for S. L. Miller, a scientist working in the field of Origin of Life. The computer, therefore, retrieves works by and citing Millers from many different fields. Further, the profile also directs a search for references having the word Venus. This yields Venus Fly-Trap literature as well as that of the planet Venus. But a better designed profile could result in prohibitive costs. In this service, the specific word, stem, etc. and how it is used in the profile determines the charge. It may be more efficient in the long run to wade through the nonrelevant matter picked up by a more general, less costly term.

The equally low precision percentage for the BIOSIS Standard Profile was expected since it is a service designed to cover the larger field of biochemical evolution. Its format gave it an advantage which allowed for optimum retrieval by the Searcher. All indexing terms assigned a reference are listed with the citation. The Searcher, therefore, did not have to rely on title alone as with other computerized output.

The abstracting character of the manually searched sources was reflected in the smaller total number of items retrieved from those sources. Chances for retrieval of only relevant articles was greatly increased. And the resultant precision values are very high. As was expected, <u>CA</u> had substantial precision and recall values when compared to the other sources. This could be a result of its broad coverage - the material was there to be found. The biggest surprise to the Searcher was the excellent performance of *IAA*. It had the highest recall value and the second highest precision value among all the sources. Since *IAA* and *STAR* are printed announcements of the RECON data-base input, the 12 percentage points difference leads the Searcher to suspect that a better designed RECON search strategy is possible. The RECON search should have retrieved, at least, the 117 items that the *IAA* retrieved.

Of the automated systems, the NASA RECON was expected to perform much better than it did. Since the mission of this information retrieval system is to serve all NASA interest groups, it should optimumly include all references of interest to this literature search, and yet, its recall value was readily surpassed by that of the commercial service, ASCA. Again, all evidence seems to point to Searcher inadequacies.

Because of its limited subject coverage and the broad subject terms which must be used in this search, MEDLARS was expected to yield few relevant references and a good deal of "noise" with a resultant low recall and precision percentages. It did just that. Because the subject area of this literature search is so much broader than the MEDLARS data-base coverage, the results are not adequate for a valid comparison with the results of studies done by Virgo,[10] Miller[11] and Lancaster.[2] Table III shows the wide discrepancies between those studies which were limited to medical questions and the present multidisciplinary study.

Table III. Comparative Results of MEDLARS Performance in Various Studies

Characteristics	Virgo	Miller	Lancaster	Present Study
Recall Percentage	50.4	48	57.7	10
Precision Percentage	54.3	18	50.4	13

UNIQUENESS AND NOVELTY CHARACTERISTICS:

The uniqueness and novelty characteristics are shown in Table IV. Uniqueness is the percentage of relevant references which were found exclusively by a source and is expressed as the ratio of the number of unique items retrieved by a system to the total relevant references found by all sources or systems. Novelty is defined as the percentage of relevant items brought to the User's attention for the first time by this literature search and is expressed as the ratio of the number of novel items retrieved from a source, or system, to the total number of relevant items retrieved by all sources or systems.

The novelty values range from CA's high of 27% to a low of 0.7% for STAR. Again, because of its very broad coverage, CA has achieved the highest novelty value. However, the ASCA yield with its 62 (22%) new relevant references may be of considerably more value to the User because of its currency factor. Although there is high duplication among the novel items, much less effort and time is required to retrieve them from ASCA.

Table IV. Uniqueness and Novelty Characteristics for Search Components

Search Component	Unique to Source	Uniqueness %	Novel to User	Novelty %
ASCA	18	6.4%	62	22%
RECON	1	0.3%	26	9%
BIOSIS Std. Prof.	7	2.5%	19	6.8%
MEDLARS	3	1.1%	20	7.1%
BA	4	1.4%	14	5%
CA	28	10%	75	27%
IAA	3	1.1%	43	15%
STAR	0	0	2	0.7%

Perhaps the most telling measure of the value of a source to a literature search of this nature is the uniqueness characteristic - how many items did that source alone contribute to the total? A quick glance at the second column figures shows that STAR has no value to the completed literature search. The three relevant items retrieved from this source were duplicated in all cases by the RECON search. The extremely low number of references unique to RECON is due mainly to the duplication of IAA and STAR coverage. Theoretically, IAA should, also, have 0% uniqueness since it is merely an announcement bulletin for published literature being indexed into the RECON data-base. Its comparatively high 15% uniqueness value is again, strongly indicative of weak search strategy with the RECON search. The generally low uniqueness values reflect the widespread duplication of the data-bases of all sources.

UNIT SEARCH TIME CHARACTERISTIC:

At the beginning of the study, time charts were kept for all searching but it soon became evident that this would not be a fair basis for comparison. It was possible to record accurately the time spent searching indexes and transferring bibliographic information to the Searcher's cards for those sources searched manually. This was not possible for all of the computer searches. Since this was a real search situation, computer printouts were scanned as received. Often the ASCA printouts were so short as to take just a couple of minutes. Fractions of a minute thus become crucial to the total time picture. Since this was also true of the monthly BIOSIS Standard Profile service and MEDLINE updates, the Searcher decided to calculate an average scanning time per citation and hopefully reduce the margin of error. This was done by recording precisely the time taken to scan a specific number of references from a source and computing an average scanning time for each citation for each source, then multiplying by the total number of references retrieved by that system. The times differed from system to system because of variation in format and data given.

Time for recording bibliographic data was treated in a like manner and an average recording time per citation was calculated. This was multiplied by the number of references recorded from that source - both relevant and subsequently nonrelevant. This unit time also varied among sources due to ease of converting from one format to that used for this study. These two factors were then summed to arrive at a total search time for that system. For example:

ASCA Search.

35 citations scanned in 5 minutes = 1/7 min per citation

1/7 x 1518 references retrieved = 217 minutes of scanning time

134 relevant or (?) relevant citations recorded

x

2 minutes recording time per reference = 268 minutes recording time

217 minutes scanning time

485 minutes search time

The total search time for each source then was the sum of the time spent scanning the published index or the computer printout for relevant citations and the time necessary to record the bibliographic data of those citations. Results of these measurements are shown in Table IV along with an extended unit of measurement for comparison.

To know that 485 minutes was spent on the ASCA search and only 114 minutes on the MEDLARS search does not really serve as a valid basis for comparison since the ASCA search yielded 109 relevant citations and the MEDLARS search only 30. In order to arrive at a common denominator for analyzing each source's time characteristic, a unit search time was computed by dividing the total search time for a source by the number of relevant references found in that source. For example:

$$\frac{\text{485 minutes search time for ASCA}}{\text{109 relevant references found in ASCA}} = \text{4.45 minutes/relevant citations}$$

Each relevant citation found in the ASCA search cost the Searcher

4.45 minutes of her time. Results for all sources are displayed in Table V.

Table V. Time Characteristics for Search Components

Component	Total Search Time - minutes	Relevant Citations	Unit Time Cost
ASCA	485	109	4.45 min/cit
RECON	260	84	3.1 min/cit
BIOSIS Std. Prof.	196	31	6.32 min/cit
MEDLARS	114	30	3.8 min/cit
BA	615	20	30.7 min/cit
CA	1878	110	17.1 min/cit
IAA	503	117	4.3 min/cit
STAR	258	3	86 min/cit

The low unit time cost of the computer searches were certainly no surprise, but the extremely competitive value for IAA was most unexpected. When compared with CA which yielded only 7 less relevant references, the difference becomes even more dramatic. Several factors may account for this. First of all, time was saved since semi-annual indexes were available for IAA but not for CA. Second, the classified arrangement of the IAA subject index greatly reduced search time. Third, the use of full titles in the IAA subject index reduced the number of abstracts to be consulted for relevancy. For example, all references indexed under a particular term by IAA are listed by full title under that heading. The term "abiogenesis" in volume 13, number 6 listed 16 items.

The titles indicated that all 16 were highly relevant. At least 6 terms must be consulted in the CA keyword subject index to retrieve abstract numbers of these same references.

The three citations retrieved from STAR were extremely costly in time. Not only did the Searcher spend 86 minutes for each of the citations, but, when the uniqueness percentages are analyzed, it is noted that all three are duplicated.

Each relevant citation from BA cost almost twice that for the ones retrieved from CA. The format of the permuted B.A.S.I.C. index is probably the biggest factor in slowing the search of BA.

The CA total search time is maximized since no cumulative subject indexes were available at the proper time. The cumulative issue did not arrive in the library until the individual issues for that volume had been searched. Time did not allow for a duplicate search when the cumulative issues did arrive. As a consequence, it was necessary to search 78 individual subject indexes.

Within the computer search mode, the BIOSIS Standard Profile search was the most costly in time. In spite of a format easily read and transcribed, the many extraneous references greatly increased the total scanning time for a source already low in precision. The RECON search achieved the best results with a time cost of 3.1 minutes per relevant citation, followed closely by the MEDLARS search with a time cost of 3.8 minutes per relevant citation. The Searcher believes this to be the result of the greater control exercised on these two searches and the subsequent smaller number of retrievals. Time was not spent with the type of "noise" that was found in the ASCA and BIOSIS Standard Profile searches.

DOLLAR COST CHARACTERISTICS:

No common denominator for costs could be ascertained for the eight sources. With the data available, absolute costs in dollars could be determined only for the two commercial computerized services, ASCA and BIOSIS Standard Profile. Because of the great variables involved, no attempt to compare machine search costs with manual costs will be made. However, there are some natural subgroupings which can be analyzed to some extent. Of the computerized sources, two are commercial current awareness ventures subscribed to by the User and administered by the library. Table VI summarizes the cost characteristics of these two services. The vast differences in the unit costs reflect the degree of personal tailoring in the profiles.

Table VI. Cost Characteristics for Commercial Computer Search Services

Commercial Service	Annual Cost	18 months Cost	Librarian Search Cost @$5/hour	Unit Cost per citation (relevant)
ASCA	$392	$588	$40.40	$5.76/cit
BIOSIS Std. Prof.	$ 50	$ 75	$16.35	$ .52/cit

The two remaining computerized sources are subsidized by the U. S. Government and at the time of this study were free to the User. The costs involved were only those of the Searcher's time. Table VII summarizes the pertinent data.

Table VII. Cost Characteristics for Subsidized Computer Searches

Subsidized Service	Librarian Search Cost	Unit Cost Per Relevant Citation
RECON	$21.50	$ .26/cit
MEDLARS	$ 9.50	$ .32/cit

Within the manual search mode, the great number of variables make any attempt to calculate a true unit cost impossible. No data on costs of acquiring, cataloging, storing or use-amortization of the printed indexes were available for this study. For strictly information purposes, yearly subscription rates and other costs factors are shown in Table VIII.

Table VIII. Costs Characteristics for Manual Search Components

Service	Annual Subscription Rate	18 months Cost	Librarian Search Cost	(Unit Cost) (Per Citation)*
BA	$1000	$1500	$ 51.25	($2.56)*
CA	$2400	$3600	$156.50	($1.42)*
IAA	$ 240	$ 360	$ 41.90	($ .36)*
STAR	$ 118 (Cumulative Indexes Extra)	$ 177	$ 21.50	($7.17)*

*These values are based on the Librarian search time only and do not take any other costs, direct or indirect into consideration.

It is interesting to note that even without considering the subscription price, the costs of the manual searches generally were higher than those in the computer mode.

MANUAL AND COMPUTER COLLECTIVE RESULTS:

Data on the collective efforts of the four sources in each mode - manual and computer - are displayed in Table IX. As was expected the machine searches retrieved a much larger number of references, yet the number of relevant items retrieved was slightly less than that retrieved by manual means. One hundred thirty three relevant references were retrieved in common.

Table IX. Manual Vs. Computerized Searching
Summary of Search Data and Characteristics

	Manual	Machine
Total citations retrieved	285	2382
Number of relevant citations	186	177
Precision	65% (90%)	7.4% (16.75%)
Recall	66%	63%
Citations new to User	115	101
Novelty	41%	36%
Items unique to search mode	53	44
Uniqueness	19%	16%
Total search time	3254 minutes	1055 minutes
Time cost per citation	17.5 min/cit	5.95 min/cit
Total Cost	* $271.15	$750.75
Dollar Cost per citation	** $1.46/cit	$4.24/cit

*Searcher's time cost only ** Based on Searcher's time cost only

The precision figures were calculated by summing the total number of relevant references retrieved by the collective efforts of the components of each mode. Virgo[13] argues that this may not produce a realistic figure since unusually high or low performance by one or more of the components of the mode can yield results atypical of the mode as a whole. For example, the gap between the lowest and the highest number of items within the manual mode varies by 133 items, yet the highest and lowest precision percentages vary by only 20 percentage points (See Table II). To allow for the unnaturally high and low performances of STAR and ASCA, precision percentages were also calculated as Virgo suggests, by averaging the precisions percentages of the components of each mode rather than totaling raw numbers. These adjusted figures are set off in parentheses.

The manual search outperforms the machine search in four of the six characteristics displayed. (We can not be sure about the dollar cost per citation since we are showing only partial costs for the manual search.) The recall factor varied little. Since the computer data-base coverage is fairly well duplicated by that of the printed indexes used for manual searching, this closeness was expected. Also expected was the superior time performance of the computer search.

The widely-spread precision percentages was not entirely unexpected when it is remembered that the manual figures are based on the number of relevant documents found or retrieved from abstracts. Lancaster[12] feels that a more accurate comparison would have resulted if all abstracts scanned were counted as retrievals. This would have affected CA retrieval figures in particular. Its keyword subject index has no standardization of terms and broader subject headings must be consulted as well as more narrow terms. This means that many abstracts may be scanned under the broader terms in order to yield a few relevant citations. It is the belief of this Searcher that the intellectual effort expended in a manual search parallels the mechanical energy expended in a machine search. If the computer were designed to exercise this same selectivity against a data-base comprised of full text or abstracts, the increased costs at this time would probably negate the increased performance.

The novelty percentage spread was only 5 points - not a dramatic difference, but of some consequence when an exhaustive search is desired. The important factor here is that 41% of the references retrieved through the manual mode were new to the User.

Value to the overall search can be pinpointed in the uniqueness factor. The manual search contributed 53 relevant items out of a total of 280 which were not duplicated by any other source. This means almost 1/5.th of the final product would be missing if the Searcher had relied on mechanical means only. Conversely, the machine search contributed 44 or almost 1/6 th of the search. From this point of view, the uniqueness results of the two modes seem somewhat more substantial.

CHAPTER 3

SUMMARY AND CONCLUSIONS

Table X was designed to summarize the results of the study by search component relationship rather than number values. Components have been ranked on a basis of 1 to 8, with 1 being the most desirable performance and 8, the least desirable. Upper positions on the chart show the most valuable components. A dotted line has been used to separate "good" performance from "bad" with the manual search components underlined. A quick glance at the concentration of underlined components indicates the relative value of the cumulative effort of manual versus machine modes.

Table X. Ranked Summary of Characteristics for Search Components

	Precision	Recall	Novelty	Uniqueness	Time Cost Per Citation	Dollar Cost Per Citation
1	<u>STAR</u>	<u>IAA</u>	<u>CA</u>	<u>CA</u>	RECON	RECON
2	<u>IAA</u>	<u>CA</u>	ASCA	ASCA	MEDLARS	MEDLARS
3	<u>BA</u>	ASCA	<u>IAA</u>	BIOSIS	<u>IAA</u>	BIOSIS
4	<u>CA</u>	RECON	RECON	<u>BA</u>	ASCA	ASCA
5	RECON	BIOSIS	MEDLARS	<u>IAA</u> & MEDLARS	BIOSIS	-
6	MEDLARS	MEDLARS	BIOSIS		<u>CA</u>	-
7	ASCA	<u>BA</u>	<u>BA</u>	RECON	<u>BA</u>	-
8	BIOSIS	<u>STAR</u>	<u>STAR</u>	<u>STAR</u>	<u>STAR</u>	-***

***Insufficient data available for rankings

To arrive at a numerical ranking of the eight services used, one to eight points were assigned on the reverse basis as position was awarded above; 8 points for the most desirable, 1 point for the least desirable. Dollar cost information was excluded since it was incomplete. For example, CA won 16 points for two first positions, 7 points for a second, 5 for a fourth and 3 for a sixth. This made a total of 31 points. When the points for each service were totaled, the following numerical rankings were established:

Service	Points
IAA	31
CA	31
ASCA	27
RECON	24
MEDLARS	21
BIOSIS Std. Prof.	18
BA	17
STAR	12

This study has reinforced the Searcher's belief that there are so many variables in every search, absolute and accurate quantification is impossible. Search strategy, data-base coverage, indexing policy and vocabulary, and index format differ with every system. Therefore, we are unable to state that any of the results are absolute or that any of these results will hold true for a future search. We can only state that for the purposes of future literature surveys on the subject of

Chemical Evolution and Origin of Life, the results of this study indicate:

1. that the most effective sources for overall recall, precision, novelty and uniqueness are Chemical Abstracts, International Aerospace Abstracts and the Automatic Subject Citation Alert (ASCA) service.
2. that the most efficient sources in terms of time and costs are the two subsidized computerized sources, RECON and MEDLARS.
3. that a more proficient search could result by
    a. discontinuing the STAR search.
    b. designing a more precise profile for the ASCA search.
    c. keeping MEDLARS service only so long as it is free.
    d. keeping BIOSIS Standard Profile service as long as the present low cost holds. Otherwise this service should be discontinued and a manual search of BIORESEARCH INDEX in conjunction with BA be initiated to replace its uniqueness value.
4. that there is a sore need for more cost information before valid conclusions can be reached concerning manual search efficiency.
5. that a better designed search strategy for RECON is needed, one proficient enough to obviate the need for IAA.

Although the manual search had a slightly better overall performance, it is concluded that neither mode is adequate for the production of a comprehensive bibliography of a multi-disciplinary nature. For an exhaustive survey of the literature, the efforts of both are necessary.

REFERENCES CITED

1. Cleverdon, Cyril W. Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems. Aslib Cranfield Research Project. Cranfield, England, Aslib, 1962.

2. Lancaster, F. W. Evaluation of the MEDLARS Demand Search Service. Washington, D. C., U. S. Public Health Service, 1968.

3. Salton, Gerard. Automatic Information Organization and Retrieval. New York, McGraw-Hill, 1968.

4. Spencer, Carol C. "Subject Searching with Science Citation Index: Preparation of a Drug Bibliography Using Chemical Abstracts, Index Medicus, and Science Citation Index 1961 and 1964," American Documentation 18(2): 87-96, April, 1967.

5. Barker, F. H., Veal, D. C., Wyatt, B. K. "Comparative Efficiency of Searching Titles, Abstracts and Index Terms in a Free-Text Data Base," Journal of Documentation 28(1): 22-36, March, 1972.

6. Angstadt, Howard P. "Searching the Current Chemical Literature by Computer," Journal of Chemical Documentation, 10(4): 277-81, 1970.

7. Lynch, J. T., Smith, G. D. W. "Scientific Information by Computer," Nature 230(5290): 153-6, March 19, 1971.

8. Searle, R. H. "Human vs. Machine Selection for Current Awareness in Mass Spectrometry," Journal of Documentation, 26(3): 221-9, September, 1970.

9. Ohta, Miwa. "A Comparison of Some Demand Subject Searches: Machine vs. Human," Medical Library Association Bulletin 55(4) : 408-15, October, 1967.

10. Virgo, Julie A. "An Evaluation of Index Medicus and MEDLARS in the field of Ophthalmology," Journal of the American Society for Information Science 21 : 254-63, July-August, 1970.

11. Miller, R. R. "A Study of Searching the Eye Research Literature," American Documentation 19 : 223-39, July, 1968.

12. Lancaster, F. W. "The Evaluation of Published Indexes and Abstract Journals," Bulletin of the Medical Library Association 59(3) : 479-94, July, 1971.

13. Virgo, Julie A. "Study of Searching the Eye Literature (Letter)," American Documentation 20(1) : 98-99, January, 1969.

BIBLIOGRAPHY

Angstadt, Howard P. "Searching the Current Chemical Literature by Computer," Journal of Chemical Documentation 10(4): 277-81, 1970.

Barker, F. H., Veal, D. C., Wyatt, B. K. "Comparative Efficiency of Searching Titles, Abstracts and Index Terms in a Free-Text Data Base," Journal of Documentation 28(1): 22-36, March, 1972.

Chenery, Peter J. "Cost of Computer Searching," RQ 12, 251-58, Spring, 1973.

Cleverdon, Cyril W. Report on the Testing and Analysis of An Investigation into the Comparative Efficiency of Indexing Systems. Aslib Cranfield Research Project. Cranfield, England, Aslib, 1962.

Duncan, E. E. Current Awareness and the Chemist. Metuchen, N. J., Scarecrow Press, Inc., 1972.

Herner, S. and Vellucci, M. J. Selected Federal Computer-Based Information Systems. Washington, D. C., Information Resources Press, 1972.

Keen, Michael. "Search Strategy Evaluations in Manual and Automated Systems," ASLIB Proceedings 20(1): 65-81, January, 1968.

Kiehlmann, Eberhard. "Organization and Efficient Manual Searching of the Major Chemical Titles and Abstract Publications," Journal of Chemical Documentation 13(2): 78-82, May, 1973.

King. D. W. and Bryant, E. C. The Evaluation of Information Services and Products. Washington, D. C., Information Resources Press, 1971.

Kochen, Manfred (ed). The Growth of Knowledge: Readings on Organization and Retrieval of Information. New York, Wiley, 1967.

Lancaster, F. W. "The Evaluation of Published Indexes and Abstract Journals," Bulletin Medical Library Association 59(3): 479-94, July, 1971.

Lancaster, F. W. Evaluation of the MEDLARS Demand Search Service. Washington, D. C. U. S. Public Health Service, 1968.

Lancaster, F. W. Information Retrieval Systems: Characteristics, Testing and Evaluation. New York, Wiley, 1968.

Loosjes, Th. P. On Documentation of Scientific Literature. London, Buttersworth and Hamden, Conn., Archon, 1967.

Lynch, J. T., Smith, G. D. W. "Scientific Information by Computer," Nature 230(5290): 153-6, March 19, 1971.

Miller, Russell R. "A Study of Searching the Eye Research Literature," American Documentation 19:223-39, July, 1968.

Ohta, Miwa. "A Comparison of Some Demand Subject Searches: Machine vs. Human," Medical Library Association Bulletin 55:408-15, October, 1967.

Project Intrex Staff; Massachusetts Institute of Technology. Project Intrex. Semiannual Activity Report, 15 March 1971 - 15 September 1971. Cambridge, Massachusetts, M.I.T., Intrex PR-12, 15 September 1971.

Salton, Gerard. Automatic Information Organization and Retrieval. New York, McGraw-Hill, 1968.

Salton, Gerard. "Evaluation Problems in Interactive Informational Retrieval," Information Storage and Retrieval 6(1): 29-44, May, 1970.

Searle, R. H. "Human vs. Machine Selection for Current Awareness in Mass Spectrometry," Journal of Documentation 26(3): 221-9, September, 1970.

Spencer, Carol C. "Subject Searching with Science Citation Index: Preparation of a Drug Bibliography Using Chemical Abstracts, Index Medicus, and Science Citation Index 1961 and 1964," American Documentation 18(2): 87-96, April 1967.

Studies in Indexing and Cataloging. Detroit, Michigan, Management Information Services, 1972.

Vickery, B. C. Information Systems. London, Buttersworth and Hamden, Connecticut, Archon, 1973.

Vickery, B. C. Techniques of Information Retrieval. London, Buttersworth and Hamden, Connecticut, Archon, 1970.

Virgo, Julie A. "An Evaluation of Index Medicus and MEDLARS in the Field of Ophthalmology," Journal of the American Society for Information Science 21:254-63, July-August, 1970.

Virgo, Julie A. "Study of Searching the Eye Literature (Letter)," American Documentation 20(1): 98-9, January, 1969.

Weisman, H. M. Information Systems, Services, and Centers. New York, Becker and Hayes, 1972.

APPENDIX A

APPENDIX A

Subject Headings Used To Retrieve Chemical Evolution and Origin of Life Literature

Abiogenesis
Biochemical Evolution
Biogenesis
Biogeochemistry
Carbonaceous Chondrites
Chemical Evolution
Exobiology
Extraterrestial Life
Origin of Life
Primitive Atmosphere, Earth, Environment or Ocean

<u>Group A</u> --- used in conjunction with --- <u>Group B</u>

Group A		Group B
Amino Acid(s)	Organic Compounds	Apollo
Ammonia	Organic Matter	Extraterrestrial Environment
Blue-green Algae	Peptides	Interstellar Molecules
Carbohydrates	Phosphorylation	Jupiter Atmosphere
Fatty Acids	Photosynthesis	Lunar Environment
Formaldehyde	Porphyrins	Mars Atmosphere
Hydrogen Cyanide	Proteins	Prebiotic
Methane	Proteinoids	Precambrian
Nicotinic Acid	Purines	Precellular
Nucleosides	Pyrimidines	Primordial
Nucleotides		Venus Atmosphere

APPENDIX B

APPENDIX B

asca

AUTOMATIC SUBJECT CITATION ALERT

a service of the **INSTITUTE FOR SCIENTIFIC INFORMATION**

5 DR. KEITH KVENVOLDEN
NASA - AMES RESEARCH CENTER
CHEMICAL EVOLUTION,239-9
MOFFETT FIELD, CALIFORNIA 94035

M1440 ACCOUNT NO.

M1440

31 JUL 73 ASCA PROFILE REVISION FORM

TEAR ON VERTICAL LINE. INSERT CARBON. CROSS OUT DELETES. ADD TERMS ON LAST SHEET USING NEXT HIGHER NUMBER. RETURN ORIGINAL.

THIS ASCA PROFILE HAS BEEN IN EFFECT 7 MONTHS

TERM NO.	NAME, INITIALS OR OTHER TERM	CITED PUBLICATION OR (CLASS OF TERM)	VOL (TYPE)	LOW PAGE	HIGH PAGE	YR	$
1	OPARIN AI	(CITED AUTHOR)					
1A	OPARIN	(CITED AUTHOR)					9
2	MILLER SL	(CITED AUTHOR)					
2A	MILLER	(CITED AUTHOR)					9
3	CRO J	(CITED AUTHOR)					
3A	DRO	(CITED AUTHOR)					9
4	FOX SW	(CITED AUTHOR)					
4A	FOX	(CITED AUTHOR)					9
5	CALVIN M	(CITED AUTHOR)					
5A	CALVIN	(CITED AUTHOR)					11
6	PONNAMPERUMA C	(CITED AUTHOR)					
6A	PONNAMPERUMA	(CITED AUTHOR)					9
22	ORGEL LE	(CITED AUTHOR)					
22A	ORGEL	(CITED AUTHOR)					13
26	SAGAN C	(CITED AUTHOR)					
26A	SAGAN	(CITED AUTHOR)					9
31	JUPITER	(WORD)	(TP 1)	DA			7
33	WALD G	ANN NY ACAD SCI	69	352		57	3
34	BERKNER LV	P NAS US	53	1215		65	3
35	ABELSON PH	P NAT ACAD US	55	1365		66	3
38	UREY H/	(CITED AUTHOR)					
38A	UREY	(CITED AUTHOR)					16
39	ANDERS E	(CITED AUTHOR)					
39A	ANDERS	(CITED AUTHOR)					9
40	KVENVOLDEN K	(CITED AUTHOR)					
40A	KVENVOLDEN	(CITED AUTHOR)					9
41	EGLINTON G	(CITED AUTHOR)					
41A	EGLINTON	(CITED AUTHOR)					9
42	SCHOPF JW	(CITED AUTHOR)					
42A	SCHOPF	(CITED AUTHOR)					9
43	CLOUD PE	(CITED AUTHOR)					
43A	CLOUD	(CITED AUTHOR)					9
44	NAGY B	(CITED AUTHOR)					
44A	NAGY	(CITED AUTHOR)					9
45	PREBIO/	(WORD)	(TP 1)	DA			7
46	CHEMICAL EVOL/	(WORD)	(TP 1)	DA			
46A	CHEMICAL -EVOL/	(WORD)	(TP 1)	DA			10
47	ABIO/	(WORD)	(TP 1)	DA			7
48	ORGANIC GEOCHEM/	(WORD)	(TP 1)	DA			
48A	ORGANIC -GEOCHEM/	(WORD)	(TP 1)	DA			10
49	METEORIT/	(WORD)	(TP 1)	DA			7
50	EXTRATERRESTRIAL	(WORD)	(TP 1)	DA			7
50A	EXTRA TERRESTRIAL	(WORD)	(TP 1)	DA			
50B	EXTRA -TERRESTRIAL	(WORD)	(TP 1)	DA			
51	PRECAMBRIAN	(WORD)	(TP 1)	DA			7
[illegible]	CARBONACEOUS CHONDRIT/	(WORD)	(TP 1)	DA			10

asca

AUTOMATIC SUBJECT CITATION ALERT

a service of the **INSTITUTE FOR SCIENTIFIC INFORMATION**

M1440 ACCOUNT NO.

31 JUL 73 ASCA PROFILE REVISION FORM

TEAR ON VERTICAL LINE. INSERT CARBON. CROSS OUT DELETES. ADD TERMS ON LAST SHEET USING NEXT HIGHER NUMBER. RETURN ORIGINAL.

THIS ASCA PROFILE HAS BEEN IN EFFECT 7 MONTHS

TERM NO.	NAME, INITIALS OR OTHER TERM	CITED PUBLICATION OR (CLASS OF TERM)	VOL (TYPE)	LOW PAGE	HIGH PAGE	YR	$
52A	CARBONACEOUS -CHONDRIT/	(WORD)	(TP 1)	DA			
53	INTERSTELLAR MOLECUL/	(WORD)	(TP 1)	DA			10
53A	INTERSTELLAR -MOLECUL/	(WORD)	(TP 1)	DA			
54	VENUS	(WORD)	(TP 1)	DA			8
55	MARS	(WORD)	(TP 1)	DA			7
56	BERNAL JD	PHYSICAL BASIS LIFE				51	4
57	HALDANE JBS	ORIGIN LIFE		242	251	67	4
57A	BERNAL JD	ORIGIN LIFE	ED	242	251	67	
58	HOLLAND HD	PETROGRAPHIC STUDIES		447	477	62	4
59	RUBEY WW	CRUST EARTH		631	650	55	4
59A	POLDERVAART A	CRUST EARTH	ED	631	650	55	
60	KAPLAN IR	(CITED AUTHOR)					9
60A	KAPLAN	(CITED AUTHOR)					
61	JOVIAN	(WORD)	(TP 1)	DA			7
62	MARTIAN	(WORD)	(TP 1)	DA			7
63	COMET/	(WORD)	(TP 1)	DA			7
64	ASTEROID	(WORD)	(TP 1)	DA			7
65	CARBON	(WORD)	(1 A1)	HB			24
66	NITROGEN	(WORD)	(1 A1)	HB			14
67	ISOTOP/	(WORD)	(1 A2)	HD			15
68	STEVENSON FJ	(CITED AUTHOR)					9
68A	STEVENSON	(CITED AUTHOR)					
69	SOWDEN FJ	(CITED AUTHOR)					9
69A	SOWDEN	(CITED AUTHOR)					
70	HUMIC	(WORD)	(TP 1)	DA			7
71	FULVIC	(WORD)	(TP 1)	DA			7
		TOTAL $ NOW IN USE					392

APPENDIX C

MEDLARS SEARCH REQUEST APPENDIX C

MEDLARS SEARCH STATION
PACIFIC SOUTHWEST
REGIONAL MEDICAL LIBRARY SERVICE
BIOMEDICAL LIBRARY
THE CENTER FOR THE HEALTH SCIENCES
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024
(213)825-5341

DATE: 5 February 1973

Is this your first request to MEDLARS? [ ] YES [X] NO

1. INDIVIDUAL WHO WILL ACTUALLY USE THIS BIBLIOGRAPHY

First	Middle	Last Name	TELEPHONE NO.
Dr. Keith		Kvenvolden	965-5181

2. TITLE

Chief, Chemical Evolution Branch

3. ORGANIZATION (Department, Bureau, Branch, Division, etc.)

Life Sciences Library, N239-13
NASA, Ames Research Center

4. ORGANIZATION (University, Corporation, Company, etc.)

Government Agency

5. STREET ADDRESS AND CITY

Moffett Field, California, 94035

6. STATE AND ZIP CODE

CALIF. 94035

7. REQUEST SUBMITTED BY (If different from above)

Linda Gufstasen, Ref. Librarian

TELEPHONE NO. 965-5387

8. SEARCH ANALYST (Leave Blank)

ALL OF THE QUESTIONS THAT FOLLOW ARE DESIGNED TO PROVIDE INFORMATION NEEDED TO DEVELOP A BIBLIOGRAPHY THAT IS RESPONSIVE TO YOUR NEEDS. YOUR CARE IN PROVIDING FULL INFORMATION WILL AFFECT THE USEFULNESS OF THE CITATIONS THAT YOU WILL RECEIVE.

9. DETAILED STATEMENT OF REQUIREMENTS: Please describe, in your own words, the subject matter for which the search is to be conducted. Be as specific as possible. Define any terms that may have special meaning in your request. Also if there are points NOT to be included, please state these.

This is a request for an update of previous searches no. 100151 and no. 10753 which together covered material published from January 1964 through March 1972. We would like to have it updated through the present date (February, 1973).

The original request read as follows:

"We desire a comprehensive search of MEDLARS literature on chemical evolution and the origin of life in the universe. Terms which might clarify in what areas we are interested are:

Origin of Life
Chemical evolution
Pre-Cambrian organic chemistry
Exobiology (life on other planets)
Spontaneous generation
Biogenesis
Abiogenesis"

We would like to add the following terms:

Prebiotic synthesis (organic compounds)
Extraterrestrial organic chemistry

DEADLINE: (Allow 4 weeks)

N 4 (Formerly PHS-4667-1)
(8)

10. SEARCH PURPOSE: Please indicate the purpose for which this search will be used (e.g., preparation of a book, book chapter, journal article, or review article; for immediate clinical application; ongoing research; prospective research; grant application; paper presented at symposium, etc.). Give specific details that will put your request into context.

ongoing bibliography

11. SEARCH LIMITATIONS: Please check all boxes that are appropriate to the scope of your request. State your needs as specifically as possible, even though we may not be able to meet these precise needs in some cases. Your replies will allow the search analyst to design a strategy that, as far as possible, will avoid types of literature that are of no interest to you.

[X] NO RESTRICTIONS

[ ] HUMAN SUBJECTS

[ ] VETERINARY MEDICINE: If only certain animals or animal groups are of interest, please list these:

[ ] ANIMAL EXPERIMENTS: If only certain animals or animal groups are of interest, please list these:

[ ] MALE

[ ] FEMALE

[ ] NORMAL STATE

[ ] CLINICAL RESEARCH *(testing of drugs or technics in humans only)*

[ ] DISEASED STATE

[ ] IN VITRO STUDIES *(of animal or human tissues or fluids only)*

[ ] CASE HISTORIES

LANGUAGE RESTRICTIONS:

[X] ACCEPT ALL LANGUAGES

[ ] ACCEPT ONLY ENGLISH

[ ] ACCEPT CERTAIN LANGUAGES ONLY *(please specify)*

AGE GROUPS: If only certain age groups are of interest, please indicate which ones:

_______ to 1 month
_______ 1–23 months
_______ 2– 5 years
_______ 6–12 years
_______ 13–18 years
_______ 19–44 years
_______ 45–64 years
_______ 65– years

GEOGRAPHIC RESTRICTIONS: If only certain regions are of interest, please list these:

12. KNOWN RELEVANT PAPERS: Please carry out a preliminary literature search of your own before submitting this request to MEDLARS, and supply full bibliographic citations below for relevant articles you have found. Wherever possible, they should be journal articles published since January 1966. These citations will be used as a guide in retrieving similar citations related to your needs. They will also be used in a later appraisal of the results of this search. If no relevant papers have been found, please state "none found".

a. Dowler, M. J., Fuller, W. D., Orgel, L. E., and Sanchez, R. A. "Prebiotic Synthesis of Propiolaldehyde and Nicotinamide", Science, 169 (3952): 1320-1321, Sept., 25, 1970.

b. Harads, K. "Chemical Evolution - Chemical Basis for Origin of Life", Protein (Tokyo) 15:855-869, July, 1970. (Japanese)

c. Fox, S. W., et al. "Bio-organic Compounds and Glassy Microparticles in Lunar Fines and Other Materials", Science 167(3918);767-770, January 30, 1970.

d. Neuman, M. W., Neuman, W. F., and Lane, K. "On the Possible Role of Crystals in the Origins of Life: IV. The Phosphorylation of Nucleotides", Curr. Mod. Biol. 3:277-283, July, 1970.

e.

If you used INDEX MEDICUS for your preliminary search please list the subject headings under which you sought citations:

ABIOGENESIS
EVOLUTION
EXTRATERRESTRIAL ENVIRONMENT
ORIGIN OF LIFE

13. SEARCH REQUIREMENTS: Please check one of the boxes below to indicate the type of search that you would prefer:

[X] A broad search designed to retrieve as many as possible of the relevant citations, but which might also retrieve many irrelevant citations.

[ ] A narrow search designed to retrieve some only of the relevant citations, but with few accompanying irrelevant citations.

NUMBER OF CITATIONS EXPECTED: Please check the appropriate box to indicate the number of journal articles dealing with the subject of your request that you consider likely to have been published since January 1966.

[ ] 0	[ ] 10 - 50	[X] 101 - 200
[ ] 1 - 9	[ ] 51 - 100	[ ] 201 - 500
		[ ] OVER 500

14. PRINT ON:

[ ] 3" × 5" cards or [X] 8½" × 11" paper

15. How did you first hear about MEDLARS?

NIH-1393-4 (Formerly PHS-4667-1)
(Rev. 10-66)

APPENDIX D

APPENDIX D

RECON Search Strategy

1. Abiogenesis
2. Biochemical Evolution
3. Biogeochemistry
4. Carbonaceous Meteorites
5. Chondrites
6. Extraterrestrial Life
7. Origins
8. Combine 1-7 (Or function)

*9. Limit to 1972-73 input

10. Amino Acids
11. Ammonia
12. Blue-green Algae
13. Carbohydrates
14. Formaldehyde
15. Methane
16. Nicotinic Acid
17. Nucleosides
18. Nucleotides
19. Organic Compounds
20. Photosynthesis
21. Porphyrins
22. Proteins
23. Proteinoids
24. Phosphorylation
25. Peptides
26. Pyrimidines
27. Purines
28. Adenines
29. Alanine
30. Cysteine
31. Lysine
32. Methonine
33. Combine 10-19 (Or function)
34. Combine 20-29 (Or function)

*35. Combine 33, 34 and 30-32 (Or function)

36. Extraterrestrial Enviornment
37. Interstellar Matter
38. Jupiter Atmosphere
39. Mars Atmosphere
40. Venus Atmosphere
41. Lunar Composition
42. Meteorites
43. Precambrain Period
44. Combine 36-43 (Or function)
45. Combine 35 and 44 (And function)
46. Limit 1972-73 input
47. Combine 9 and 46 (Or function)
48. Print 47