REMOTE EVALUATION OF A REMOTE CONSOLE INFORMATION RETRIEVAL SYSTEM (NASA/RECON)

Victor L. Coles
National Aeronautics and Space Administration

Paper for presentation at the Forum on Interactive Bibliographic Systems, to be held October 5, 1971 at the National Bureau of Standards, Washington, D.C.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

(NASA-TM-X-67544) REMOTE EVALUATION OF REMOTE CONSOLE INFORMATION RETRIEVAL SYSTEM (NASA/RECON) V.L. Coles (NASA) Sep. 1971 29 p CSCL 05B

N72-18987

FACILITY FORM 602 (ACCESSION NUMBER) (CATEGORY) G3/34 Unclas 18313

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546 SEPTEMBER 1971

REMOTE EVALUATION OF A REMOTE CONSOLE INFORMATION RETRIEVAL SYSTEM (NASA/RECON)

by Victor L. Coles
National Aeronautics and Space Administration

ABSTRACT

NASA constantly seeks evaluation from its users of a nationwide literature search service based on an interactive information retrieval system. This report explains the technique, which consists of sending out an evaluation form with each literature search, and the results derived from a compilation of the user's responses. In an eleven-month period in which evaluation forms went out with 3,001 searches, 33.6% of the forms were completed and returned. The returns showed that 88.5% of the respondents found the searches suitable to their needs, 81% learned of valuable new references from the searches, and 93.5% received the searches in time to meet their needs. The significance of relevance or precision ratio in relation to user satisfaction is discussed, and an extrapolation from the users' responses resulted in a relevance ratio of 49.3%. Some of the general comments found in the responses are analyzed as indicators of what the user s expected from the information retrieval service.

REMOTE EVALUATION OF A REMOTE CONSOLE INFORMATION RETRIEVAL SYSTEM (NASA/RECON)

by Victor L. Coles *
National Aeronautics and Space Administration

I. INTRODUCTION

To make it possible for an information retrieval system, which is growing as rapidly as NASA's, to meet the need of its users, who have a wide variety of interests and needs, NASA must get user evaluation constantly. In the words of Hoshovsky and Massey, "Information Economics is a user-oriented discipline. Its perspective is inherently that of the user since the value of data in application, the essence of information in the sense we use the term, is a function of the user's problems and the alternative knowledge sources open to him." (1)

The principal advantage of an interactive system is that the machine's rapid response to the user's manipulation of the console keyboard gives the user the opportunity of evaluating the results of his search immediately. If the bibliographic references the user receives from the system do not satisfy him, he may amend or alter his search strategy to produce a new set of results. He may resort to a browsing technique, one of which is described in detail in a recent paper by J. H. Williams, Jr., in which he states: "The purpose of browsing in a text retrieval system is to reduce the number of false hits and increase the number of true hits.... The problem is: relevant

---

* Mr. Coles is Analysis and Review Officer, Information Services Division Scientific and Technical Information Office, NASA, Washington, D.C. 20546

documents are known to exist in the data base but they were not retrieved with the original formulation of the query. The primary reason for missing documents is caused by authors employing different terms to express the concept than the searcher uses to express his query. The searcher therefore needs to perform a preliminary search through whatever material is available to recall various terms for expressing the same concept." (2) Sometimes, even though the results the user receives are not precisely within the narrow limits of the subject he originally started out to search, the information content of the retrieved items may be so valuable to him that the user declares the search a success and terminates his searching efforts.

The performance evaluation that I am about to discuss is of a different nature, one a bit more severe than that which I have just described. This is the evaluation by a requester of a printed literature search prepared for him by a search analyst seated at the console of an interactive retrieval system which may be remote from the requester. In this case, the search analyst develops his search strategy on the basis of a written search request or a discussion of a written request with the request writer. The analyst makes a decision to accept or reject the results objectively by comparison with the request statement, but without the benefit of the same specialized knowledge, education, or experience as the user . Once the search analyst terminates the searching operation and transmits the printed results to the requester, he no longer has the option of amending the search strategy to improve it. The requester, having received the search results, evaluates them

without considering the steps in the procedure that were used to obtain them. His evaluation is based on how well the results answer his information need, which may be quite different in one way or another from the written query in his search request.

II. THE EVALUATION TECHNIQUE

The NASA Scientific and Technical Information Facility (hereafter called the Facility) receives requests for literature searches from employees of NASA, NASA contractors, other U.S. Government agencies, and university libraries registered with NASA for such service. The Facility search analysts are authorized to discuss ambiguous or complex requests with the requesters. These analysts perform searches using terminals of NASA/RECON, NASA's interactive retrieval system, on a high-volume production basis. The results are printed out at night, off-line, when time is available for lengthy printouts. The analysts receive the printouts of their searches the next morning. A brief review is made of each search to see that there are no major flaws in it, but it is not edited citation by citation for content. The printout is assembled with explanatory material and mailed to its requester.

Each search mailed is accompanied by a return-addressed, franked evaluation form. The requester fills in the form with his appraisal of the search results and sends the completed form to NASA Headquarters for review. It is then forwarded to the Facility for direct feedback to the analyst. The Facility may take corrective action if it is shown to be necessary. Regular statistical reports are prepared on the answers to the evaluation form questions. An analysis of an eleven-month cumulation of evaluation responses follows.

III. WHAT THE USERS REPORTED

After a few years of development and experimental use (3), the present configuration of NASA/RECON was declared to be fully operational for the routine production of literature searches in July 1970. The results presented here are for user evaluations of searches completed from August 1970 through June 1971. In that period, the Facility mailed (SLIDE 1) out 3,001 literature searches, each accompanied by an evaluation form; and 1,015 of the forms were filled out and returned to NASA. This amounts to a 33.6% return. Hereafter, the users who returned completed evaluation forms will be called respondents.

In the returns, 88.5% of the respondents said that the search was suitable to their needs; 9.5% said it was not; and 2% left the question unanswered. Without knowing the opinion of the respondents who did not answer this question, these figures indicate that we had less than a 12% failure rate based on this question alone. (SLIDE 2 / OVERLAY 1)

Since this is only one of many services offered by NASA's Scientific and Technical Information Office (4), we also wanted to find out if these searches were only repackaging citations of which the requester had already been informed through other means or whether they had a worthwhile payoff that was unique to the literature searches themselves. In response to the question "Did the search provide any valuable new references?", 81% of the responses were "YES," 11.5% were "NO," and (OVERLAY 2) 7.5% gave no answer to the question. This did give some assurance that the users were being informed of the existence of some documents through literature searches that their other sources had not yet brought to their

attention. This was reassuring in the light of William T. Knox's reminder that "An information service competes with the individual's own sources of information." (5)

IV. RESPONSE TIME

I mentioned at the start that a primary feature of an interactive retrieval system is its quick response time. We would not want this feature to be lost to our outside requesters, even though the Facility is working regularly against a backlog of written requests. Our contract permits the Facility a maximum of five working days in which to process a literature search request in-house. Even adding on time for slow mail delivery, requesters still could receive their searches in about two weeks from the date that they mailed their requests. Is this fast enough? In answer to the question "Did you receive the search in time to meet your needs?" 93.5% said they did and only one half of one per cent said that they did not. Six per cent didn't answer the question. (SLIDE 3) (OVERLAY 1)

V. RELEVANCE

Now let's get down to the fundamental issue: the evaluation of relevance. Although Saracevic noted that relevance judgment has associated with it some remarkable regularity patterns (6), the significance of relevance, in fact its very meaning, has been questioned for many years (7, 8, 9). Nevertheless, since the system is designed to retrieve citations relevant to a given search requirement, relevance is one measure of system effectiveness.

In attempting to consolidate into a single question an inquiry into the user's evaluation of relevance, on one hand, we offered the user wide latitude in which to define relevance by suggesting that he include in his selection of relevant references those that are "either directly or generally pertinent." (SLIDE 4

On the other hand, the question ends on a severely restrictive note. In judging what is pertinent "to his requirements" (in the words of the question on our evaluation form), the user was prone to measure the relevance of the references he received by their ability to provide a finite answer to a problem he encountered in his work, rather than against the phrasing of his request as he had written it. Thus his frame of reference for quantifying relevance might be quite different from the specification of the problem entering the retrieval system.

Although the form requested a measure of relevance expressed as percentage of citations, we did not average the percentages received. Instead, the respondents' percentages of relevance were converted to the equivalent numbers of pertinent citations in each search; the number of pertinent

citations was cumulated for eleven months and then divided by the cumulated number of total citations retrieved and mailed out in that time, to get an overall relevance percentage. The 1,015 searches in the reporting period had contained 147,649 citations, of which 72,820 were judged to be relevant by the users. This results in a relevance or precision percentage of 49.3%. The average number of relevant citations per search was 72.

(OVERLAY 1)

Comparisons with studies of other systems can seldom be made in truly equivalent units. The user's estimates of relevance are affected by changes in the wording of the question from one study to another. Also, in different information systems, the operating factors that affect the number of relevant references furnished to the requester vary. For these reasons, the quantitative results obtained in evaluations of different systems do not really correspond in meaningful ways.

However, as long as at least one other large-scale study exists, comparison, albeit a superficial one, is inevitable. When Lancaster (10) made his two-year study of the MEDLARS search system, he found that the precision ratio for that system varied, depending on the mode of interaction between the requester, his local librarian, and/or the MEDLARS search analyst, from 46.9% to 53.9%, with an overall average of 50.4%. In 109 cases, the MEDLARS analyst had discussed the search request with the requester before performing the search, as NASA Facility analysts often do, and for this sub-set the average MEDLARS precision ratio was 49.3%.

(OVERLAY 2)

VI. THE SIGNIFICANCE OF RELEVANCE IN USER SATISFACTION

The significance of a user's rating of relevance needs to be judged case by case in conjunction with the other answers the respondent provides in evaluating the search. O'Connor (and others) mentioned that the volume of documents required to meet a user's information need varies: "Does the user want any one S-document (to answer a question), a few (to start on a subject), most in the collection (an exhaustiveness needed for scientific, military, safety, or legal purposes)?" (11)

For those who wanted "a few documents (to start on a subject)," our average of 72 relevant citations per search was probably sufficient.

A. F. Goodman (12) states that, although 41% of the literature search requesters he interviewed said that they wanted "all available material," another 30% answered that one report or document would suffice. If the right document is found, the one which contains the needed answers, the relevance of the remaining citations in a search may be of little significance, no matter how numerous they are. Following this line of reasoning, we noted that although 326 users had reported that 20% or less of the citations they had received were pertinent to their (SLIDE 5) requirements, 72% of these users said that the search was suitable to their needs, and an equal percentage of them learned of valuable new references from these searches.

VII. INCLUSIVENESS

The requesters were not asked to make precise measurements of recall. As a general indicator of the inclusiveness of the search, or the thoroughness of coverage, the evaluation form asked "Was the subject adequately covered by this search?" and "Do you know specific references that should have been included?" In response to the first question, 76.5% of the respondents answered that the subject had been adequately covered; 18% answered "NO", and 5.5% did not answer. Only 14% could cite specific references that had been omitted; many of the documents that the requesters cited in response to this question were not in our collection at the time of the search, and some were not within the scope of our collection. Sixty-five per cent (65%) did not know of any documents that had been missed, and 21% did not answer this question. A more direct calculation of recall was made in-house at NASA Headquarters, which will be discussed next.

(SLIDE 6)

(OVERLAY 1)

B

VIII. IN-HOUSE EVALUATION

As user evaluations are subjective in nature, an occasional spot-check of system effectiveness is made each month by evaluating a few searches in-house at NASA Headquarters. Such searches are rated for relevance and recall. For this purpose, a citation is considered to be relevant if the title or Notation of Content contains words related to at least two of the concepts contained in the original request, and these words are in the proper syntactical relationship. Recall is based either on a manual search of printed reference tools or on a NASA/RECON dump of a few broad subject terms.

Fourteen selected searches resulted in an average recall of 40.6%. Of the 3,375 references contained in these searches, 1,815 were considered relevant, according to the definition given earlier, for a precision ratio of 53.7%.

IX. USER COMMENTS

Space was provided on the NASA evaluation form for general comments, but only half of the respondents used the space. Most of the user comments dealt with specific aspects of the subjects covered by the searches; some of these comments indicate how difficult it is for any practical system to meet user expectations.

One respondent wrote: "Only four pertinent references were listed. I already had two of them. All four contain only one of the several methods. I was interested in original, unknown methods." Another wrote: "Insufficient number of novel processing techniques were presented." In these two cases, the search analyst could not be expected to know which documents the requester already had, nor what methods the requester considered to be "novel" or "original." This kind of determination can only be made by the user.

A search was requested on the subject "Techniques for mixing powders in liquids..." but when the requester received the search he wrote: "Equipment available for mixing was desired." Had he expressed his information need in that fashion, a different search startegy might have been used.

On a search containing 450 citations, the requester commented: "Although only 40% of the material directly applied to my immediate problem, it will serve as a valuable source to colleagues in related areas. Placed in permanent file." High recall with low relevance may be helpful to an organization with diversified interests in a particular field.

One user wrote: "Pleased with the foreign material that I might have missed," but another commented: "All relevant references were in Russian and hence of no use to me." A system designed for an international collection probably should have either a language searching capability or at least the ability to limit the output to English-language documents. It is through the expression of the user's needs that the system can be kept user-oriented.

Many of the more general comments were terse words of praise; 14 included the words "very helpful"; 10 wrote "very useful"; 4 said "well done"; 2 even said "very well done.". The word "good" was the rating in 14 responses, "very good" in 9 and "extremely good" once, as well as one "superb." Twenty-four respondents rated the search "excellent" and though this rating was gratifying, it could not overshadow the comments of three different organizations that wrote "Best literature search ever received!"

X. SUMMARY AND CONCLUSIONS

To meet the changing needs of information users who have a variety of interests, constant evaluation of an information retrieval system is necessary. This study has shown that a satisfactory response for the evaluation of an interactive retrieval system may be obtained from remote users by furnishing an evaluation form with each printout of a literature search mailed to a user. NASA obtained a 33.6% return in an eleven-month period in this manner.

The results of the NASA evaluation of NASA/RECON output indicated that 88.5% of those who responded found the searches suitable for their needs and 81% learned of valuable new references from their searches. A maximum processing time of five working days, with the time for mailing the request and the finished product added on, provided sufficiently rapid service for 93.5% of the respondents.

Extrapolation from the responses on the evaluation forms indicated that the searches average 49.3% relevance (or precision), which matches the results of another large-scale computerized literature search service. A separate spot-check of recall conducted at NASA Headquarters suggested that the average recall is in the neighborhood of 46.1%.

Although general comments on the system by remote users are judged with consideration for the functions the system was designed to perform, the comments give valuable insight into the user's changing needs, and may provide worthwhile suggestions for needed system modifications.

XI. ACKNOWLEDGEMENTS

I wish to acknowledge gratefully the efforts of Mr. Philip Eckert and Mr. William Neely of the NASA Scientific and Technical Information Facility for compiling the statistics on which this report is based, I also wish to thank Miss Kay Vogelewede, Mr. C. W. Hargrave and Mr. Van A. Wente for their critical comments during the preparation of this paper.

REFERENCES

1. Hoshovsky, Alexander G., and Robert J. Massey. Information Science: Its Ends, Means and Opportunities. IN Proceedings of the American Society for Information Science, Vol. 5, Information Transfer. Annual Meeting, Columbus, Ohio, October 20-24, 1968. p50.

2. Williams. J. H.;Jr., An Iterative Browsing Technique. Annual Progress Report, Contract N0001h-70-C-0297, April 1971. International Business Machines Corporation, Federal Systems Division, Gaithersburg, Maryland AD 722672.

3. Wente, Van A., NASA/RECON and User Interface Considerations. National Aeronautics and Space Administration, Washington, D. C. Presented at the Workshop "The User Interface for Interactive Search of Bibliographic Data Bases" sponsored by the AFIPS Information System Committee. January 14-15, 1971.

4. The NASA Scientific and Technical Information System...And How to Use It. National Aeronautics and Space Administration, Washington, D. C. 20546

5. Knox, William T., How Much Information Service? IN Journal of Chemical Documentation, Vol. 2 No. 1, January 1962, p 27.

6. Saracevic, Tefko. Ten Years of Relevance Experimentation: A Summary and Synthesis of Conclusions. IN Proceedings of the American Society for Information Science, Vol. 7, The Information Concious Society, 33rd Annual Meeting, Philadelphia, Pa., October 11-15, 1970. p 35.

7. Doyle, Lauren B. Is Relevance an Adequate Criterion in Retrieval System Evaluation? IN Automation and Scientific Communication, October 1963.

8. Cuadra, C. A., and R. V. Katter. Experimental Studies of Relevance Judgments. Final Report, 3 volumes. Systems Development Corp.,Santa Monica, California. 1967. 498 p

9. Cooper, W. S. A Definition of Relevance for Information Retrieval. IN Information Storage and Retrieval, Vol. 7 pp 19-37. 1971.

10. Lancaster, F. W., MEDLARS: Report on the Evaluation of its Operating Efficiency. IN American Documentation, Vol 20, No. 2 April 1969 p 119.

11. O'Connor, John. Mechanized Indexing Methods and their Testing. IN Journal of the ACM, Vol. 11 No. 4, October 1964

12. Goodman, A. F. Flow of Scientific and Technical Information: The Results of a Recent Major Investigation. McDonnell Douglas Astronautics Company. Douglas Paper 4516 Revised April 1969. Presented to the 14th International Meeting, The Institute of Management Sciences, Mexico City, Mexico, 22-26 August 1967.

NASA LITERATURE SEARCH EVALUATION FORM

PART I (FOR THE REQUESTER OF THE SEARCH)

We would greatly appreciate your help in evaluating the work we have performed in response to your literature search request. When you or someone in your organization has had an opportunity to review the enclosed search, we would be grateful if the reviewer would answer the following questions, fold and staple the form and mail it to the address printed on the back.

1. Was the literature search suitable for your needs? Yes__; No__.

2. Of all of the references in the search, what percentage was either directly or generally pertinent to your requirements? ____%

3. Was the subject adequately covered by this search? Yes__; No__.

   If not:
   a) Was the overall scope too broad or too narrow? Broad __; Narrow __.
   b) Were the individual references too general or too specific? General__;Specific__.
   c) Were desired aspects of the subject missing? If yes, please explain in Comments (Item 7). Yes__; No__.

4. Did you receive the search in time to meet your needs? Yes__; No__.

5. Did the search provide any valuable new references? Yes__; No__.

6. Do you know specific references that should have been included? Yes__; No__.
   If yes, please identify them on the back of this form and check here to indicate that references are listed there. Listed ____.

7. COMMENTS: _____

_____

_____ (Signature) _____ (Date)

PART II (TO BE FILLED IN AT THE NASA SCIENTIFIC AND TECHNICAL INFORMATION FACILITY)

Literature Search Number _____; Number of Citations _____

Search Title _____

Scope Statement _____

Name and Address of Requester _____

_____

MAAL Registration Number _____ Requester Profile _____

PART III (TO BE FILLED IN AFTER THE COMPLETED FORM IS RETURNED TO THE NASA FACILITY)

ACTION TAKEN AS A RESULT OF THIS EVALUATION: _____

_____

_____

STATISTICAL ANALYSIS OF COMPLETED LITERATURE SEARCH EVALUATION FORMS

Report Period: August 1970 - June 1971

A. General Data

1. Number of forms returned: 1,015
2. Number of searches performed: 3,001
3. Percentage returned: 33.6%

B. Question Responses (Numbering same as on Evaluation Form)

1. Was the literature search suitable for your needs?

Yes - 88.5% (895 responses). No - 9.5% (96). Unanswered - 2.0% (24).

3. Was the subject adequately covered by this search?

Yes - 76.5% (778 responses). No - 18.0% (180). Unanswered - 5.5% (57)

4. Did you receive the search in time to meet your needs?

Yes - 93.5% (949 responses). No - 0.5% (8). Unanswered - 6.0% (58).

5. Did the search provide any valuable new references?

Yes - 81.0% (825 responses). No - 11.5% (117). Unanswered - 7.5% (73)

6. Do you know specific references that should have been included?

Yes - 14.0% (144 responses). No - 65.0% (662). Unanswered - 21.0% (20

7. Comments.

Furnished comments: 51.0% (519). Left blank: 49.0% (496)

C. Citation Acceptance Table

Total Citations: 147,649

Pertinent Citations: 72,820

Citation Acceptance Ratio: 49.3%

IN-HOUSE EVALUATIONS SEPT 1970-JULY 1971

SEARCH #	NO. OF TOTAL CITATIONS	PRECISION %	NO. OF PERTINENT CITATIONS	RECALL %	POTENTIAL PERTINENT ITEMS	
12975	35	66	23	20	115	9/3/70
13414	303	50	152	60	253	10/23/70
13415	659	32	211	80	264	10/26/70
13820	44	80	35	47	74	12/17/70
14151	107	76	81	67	121	2/5/71
14323	951	40	380	68	544	2/19/71
14487	86	70	60	75	80	3/5/71
14602	204	74	151	8.33	1813	3/18/71
14713	131	55	72	71.4	100	3/29/71
14924	314	66.7	210	48.6	486	4/16/71
14941	378	81	306	100	306	4/16/71
14980	9	66.7	6	25	24	4/22/71
15242	108	78	84	40	210	5/14/71
15700	46	96	44	60	73	7/2/71
	3375		1815		4463	

$$\frac{\text{NUMBER OF PERTINENT CITATIONS}}{\text{NUMBER OF CITATIONS RETRIEVED}} = \frac{1815}{3375} = 53.7\% \text{ PRECISION}$$

$$\frac{\text{NUMBER OF PERTINENT CITATIONS RETRIEVED}}{\text{POTENTIAL PERTINENT CITATIONS IN COLLECTION}} = \frac{1815}{4463} = 40.6\% \text{ RECALL}$$

NUMBER 1

STATISTICAL ANALYSIS OF COMPLETED LITERATURE SEARCH EVALUATION FORMS

GENERAL DATA

REPORTING PERIOD: AUGUST 1970 - JUNE 1971

NUMBER OF FORMS RETURNED: 1,015

NUMBER OF SEARCHES PERFORMED: 3,001

PERCENTAGE RETURNED: 33.6%

NASA HQ KS72-15152 (1) 9-20-71

NUMBER 2
OVERLAYS 1&2

WAS THE LITERATURE SEARCH SUITABLE FOR YOUR NEEDS?

YES 88.5%
NO 9.5%
Unanswered 2.0%

DID THE SEARCH PROVIDE ANY VALUABLE NEW REFERENCES?

YES 81.0%
NO 11.5%
Unanswered 7.5%

NUMBER 3
OVERLAY- 1

DID YOU RECEIVE THE SEARCH IN TIME TO MEET YOUR NEEDS?

YES	93.5%
NO	0.5%
Unanswered	6.0%

NUMBER 4 OVERLAYS 1&2

OF ALL THE REFERENCES IN THE SEARCH, WHAT PERCENTAGE WAS EITHER DIRECTLY OR GENERALLY PERTINENT TO YOUR REQUIREMENTS?

PERTINENT CITATIONS:	72,820
TOTAL CITATIONS FURNISHED:	147,649
CITATION ACCEPTANCE RATIO	49.3%

AVERAGE NUMBER OF PERTINENT CITATIONS PER SEARCH: 72

MEDLARS PERFORMANCE FIGURES

PERSONAL INTERACTION, 109 SEARCHES:	PRECISION RATIO - 49.3%
OVERALL AVERAGE OF 299 SEARCHES:	PRECISION RATIO - 50.4%

NASA HQ KS72 15155 (1) 9 20 71

NUMBER 5
(NO OVERLAYS)

STUDY OF 326 SEARCHES IN WHICH USER EVALUATED PERTINENCE BETWEEN 0 AND 20%

WAS THIS LITERATURE SEARCH SUITABLE FOR YOUR NEEDS?

YES 72.5%

NO 24.5%

Unanswered 3.0%

DID THE SEARCH PROVIDE ANY VALUABLE NEW REFERENCES?

YES 72.0%

NO 23.0%

Unanswered 5.0 %

NASA HQ KS72-15156 (1) 9-20-71

NUMBER 6
OVERLAY 1

WAS THE SUBJECT ADEQUATELY COVERED BY THIS SEARCH?

YES 76.5%
NO 18.0%
Unanswered 5.5%

DO YOU KNOW SPECIFIC REFERENCES THAT SHOULD HAVE BEEN INCLUDED?

YES 14.0%
NO 65.0%
Unanswered 21.0%