



EVALUATION of the MEDLARS DEMAND SEARCH SERVICE

January 1968

F.W. Lancaster
Deputy Chief, Bibliographic Services Division
National Library of Medicine

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service

"IN ORDER TO SURVIVE, A SYSTEM MUST MONITOR ITSELF, EVALUATE ITS PERFORMANCE, AND UPGRADE IT WHEREVER POSSIBLE."

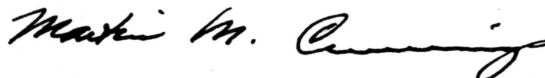
* From report GER 12760 Measures of Effectiveness and Criteria for Evaluation of a Document Processing System. Rome Air Development Center, 15 July 1966.

P R E F A C E

This report presents the results of a detailed analysis by the National Library of Medicine of the performance of MEDLARS (its Medical Literature Analysis and Retrieval System) in relation to 300 actual requests made to the system in 1966 and 1967. Thanks to this study, we now know considerably more about the requirements of MEDLARS users, how well the system is meeting these requirements, and what we must do to improve the overall performance level. The investigation is timely: the Library is now planning a second-generation MEDLARS, and the design of MEDLARS II should benefit greatly from our knowledge of factors affecting the performance of the present system.

Since this is the first large-scale evaluation of a major operating information system, and because of the diversity of subject areas covered by the study, it should be of considerable interest to the scientific community at large. Some readers, of course, may wish to take exception to parts of the methodology of the study or even view some of the analyses with reservation. In an effort to make the study as objective as possible, the design and results were reviewed by a distinguished outside advisory committee to whom we are most grateful.

To remain responsive to the demands of its users, a large scientific or technical information system must examine itself critically. We hope that a major benefit of this investigation will be the establishment of a program for the continuous quality control of MEDLARS products and services.



Martin M. Cummings, M.D.
Director
National Library of Medicine

CONTENTS

	<u>Page</u>
PART 1 DESIGN AND EXECUTION OF THE EVALUATION PROGRAM	
Introduction	8
MEDLARS: general background	9
Objectives of the test program	12
Test design: general considerations	15
Selection of user groups	15
Establishing the performance figures	19
The pretest	26
Procedures used in the conduct of the test	27
Derivation of performance figures	30
Analysis of reasons for search failures	32
 PART 2 THE TEST RESULTS	
Overall performance figures	35
The individual ratios	38
Average MEDLARS performance for the test requests	40
Analysis of causes of recall and precision failures	46
Analyses of failures: explanatory notes	46
<u>Recall and precision failures attributable to the indexing subsystem</u>	51
Failures due to exhaustive indexing or to lack of exhaustivity	54
Effect of exhaustivity levels	58
Failures due to lack of specificity in indexing	61
General observations on exhaustivity and specificity of indexing within MEDLARS	62
Evaluation of indexing as part of the input subsystem	64

<u>Recall and precision failures attributable to the searching subsystem</u>	65
Recall losses resulting from failure to cover all reasonable approaches to retrieval	65
Recall and precision failures due to variations in exhaustivity of the formulation	67
Recall and precision failures due to variations in specificity of the formulation	70
Use of "weighted" index terms	76
Other causes of searching failures	79
Joint causes of system failures	81
Effect of the 6-5-4 levels on recall and precision figures	81
<u>Recall and precision failures attributable to the index language</u>	82
Failures due to false coordinations and incorrect term relationships	91
General observations on the MEDLARS index language	98
The relationship between indexing, index language, and searching	100
<u>Recall and precision failures attributable to the area of user-system interaction</u>	101
Improving request statements	117
Experimentation with modes of interaction	125
Recall and precision failures attributable to computer processing	127
The novelty ratio	128
Factors affecting performance of a MEDLARS search	130
Variations in performance between five MEDLARS centers	145
MEDLARS indexing coverage	157
Foreign literature usage factors	159

	Journal usage factors in MEDLARS	165
	Effect of MEDLARS response time	174
	The serendipity value of MEDLARS searches	175
	Output screening	176
	Indexer consistency	180
	Requests rejected by MEDLARS	183
✓	PART 3 CONCLUSIONS AND RECOMMENDATIONS	
	Overall MEDLARS performance	185
	Upgrading the performance of MEDLARS	191
	User-system interaction	193
	The MEDLARS index language	193
	The MEDLARS searching strategies	195
×	The MEDLARS indexing	197
	Computer processing	199
	The relationship between indexing, searching and <u>MeSH</u>	200
	Use of foreign language material in MEDLARS	200
	The search printout as a content indicator	201
	Continuous quality control of the MEDLARS operation	201
	Future use of the MEDLARS test corpus	202
	PART 4 APPENDICES	
	Appendix 1. Samples of MEDLARS vocabulary, sample request and search form- ulation, sample printout	206
	Appendix 2. Specimens of forms used in evalua- tion	213
	Appendix 3. Specimen analysis worksheet	218

Appendix 4. Complete set of recall and precision ratios	230
Appendix 5. Analysis of the specificity of search formulations	249
Appendix 6. Analysis of the effect of sub-headings on false coordinations and incorrect term relationships	264