

ASLIB CRANFIELD RESEARCH PROJECT

FACTORS DETERMINING THE PERFORMANCE
OF INDEXING SYSTEMS
VOLUME I. DESIGN

by

Cyril Cleverdon, Jack Mills and Michael Keen

Part 1. Text

(A) An investigation supported by a grant to Aslib
by the National Science Foundation

Cranfield

1966

ACKNOWLEDGMENTS

The project reported in this volume followed on from the first Aslib-Cranfield project, and was again financed by a grant from the National Science Foundation. The support of the Foundation was, as usual, not restricted to financial assistance, and I am most grateful for the advice and encouragement which we received, particularly from Mrs. Helen Brownson.

The full-time members of the project group were Mr. Jack Mills, from September 1962 to August 1965, Mr. Wilfrid Lancaster in the year of 1963, and Mr. Michael Keen, who commenced in January 1963 and is still engaged on the final stages of the project. In addition, some sixty-three other persons have worked part-time at some stage. To all these people, I have to express my appreciation for their efforts.

An essential requirement of the project involved co-operation of a large number of research scientists. The response to our request was most satisfactory, and I acknowledge with thanks the generous assistance of some two hundred scientists, many of whom are known to me only by name.

As before, Aslib administered the grant and also, on this occasion made accommodation available in their headquarters in London, and I am grateful for the help given by the Director, Mr. Leslie Wilson, and many members of his staff.

I would also express my appreciation to the Principal and Senate of the College of Aeronautics for agreeing to my taking part in this project while continuing my normal duties.

Finally, there are many friends and colleagues with whom, during the past three years, I have had the opportunity of discussing the Aslib-Cranfield projects. Their comments and suggestions have always been helpful, and I am most grateful for the interest which they have shown.

CONTENTS

PART 1

		Page
Chapter 1	General considerations	1
Chapter 2	Test design	9
Chapter 3	Documents and questions	19
Chapter 4	Indexing procedures	40
Chapter 5	Formation of index languages	58
Chapter 6	Testing techniques	90
Chapter 7	Additional tests	106
Chapter 8	Comments	113
	References	118

Part 2 Appendices

Index

LIST OF TABLES AND FIGURES

		Page
Figure 2.1	N.A.S.A. search system analysis sheet	12
Table 3.1	Bibliographical origin of base documents used in the test	20
3.2	Country of residence of authors of base papers	20
3.3	Comparison of authors' country of residence and country of publication	20
3.4	Relevance assessments of documents as decided by authors in relation to their search questions	25
3.5	Grades of relevance as decided by the authors	25
3.6	Relevance assessments giving a comparison of basic and supplementary questions	26
3.7	Relevance assessments giving a comparison of basic and supplementary questions for all grades of relevance	27
3.8	Breakdown of 312 documents retrieved by bibliographic coupling at strength of 7 or more	35
3.9	Examples of question/title matches for relevant documents	35
3.10	Relevance grades of documents with specified question/title match	37
3.11	Comparison of the cited and additional documents with specified question/title match	37
3.12	Comparison of recall performance of relevant cited and additional documents in relation to 25 questions	39
Figure 4.1	Indexing sheet for Document 1590	51
5.1	Natural language single term data	59
5.2	Patterns of term usage	60
5.3	Sample sheet from schedules of single terms	72

		Page
Figure 6.1	Master indexing sheet for Document 2076	92
6.2	Posting sheet for 'FLOW' in relation to Documents 1745-2116	94
6.3	Starting term authority sheet showing terms related to 'FLOW'	95
6.4	The 'Beehive' filing cabinet	95
6.5	Search starting terms for Question 181	95
6.6	Search sheet for Question 181 in relation to Documents 1931-1992	96
6.7	Score sheet for Question 181 in relation to Documents 1956-1992	99
6.8	Results sheet for Question 181 for index languages 1 to 6	101
6.9	Processing of Document 2076 in relation to Question 51 for analysis of interfixing and partitioning	103
6.10	Score sheet for links with Document 2076 for Question 51	104
6.11	Instruction sheet for search with controlled term vocabulary	104
7.1	Citation index sheet	108
7.2	Citation index reference card	109
7.3	Citation index master card	109
7.4	Bibliographic coupling card	111
7.5	Score sheet for bibliographic coupling for Question 34	111
7.6	Recalculated bibliographic coupling card	111