CHAPTER 7

CITATION INDEXING AND BIBLIOGRAPHIC COUPLING

> It is true that librarians do an almost religious job of storing information; it is placed on record, but without evaluation, and much of it is not worth its rental of space. Each operational unit tends to use its own special language, and translators are very few. In warfare, the questioning of returning forces is regarded as a highly skilled speciality. Would social action (and perhaps social research) not gain much if an analogous speciality could be created to assess and consolidate relevant information? Such an organization could not confine its attention to information retrieval, no matter how efficient such a retrieval process might be. Storage and retrieval systems do not represent evaluation and consolidation of information .
>
> L.T. Wilkins: Social Deviance Page 111

In Chapter 7 of Vol. I, an account was given of the compilation of a citation index, and the subsequent preparation of bibliographic coupling groups, with a view to an evaluation being made of this index.

As stated in this earlier account, it is a matter for some argument as to how this type of index can be tested in an experimental situation. In carrying out a test in an operational environment, there would be no difficulty beyond the effort required, but although several different ways of presenting the results of this test have been considered, there does not appear to be any procedure which can be considered entirely satisfactory. For this reason two sets of figures are being given; the first method probably results in a performance which is better (in comparison with the results obtained with conventional systems) than should be the case, but it has the major advantage that it does not involve any manipulation of the test results and therefore permits direct comparison to be made between different subsets of questions, different relevance decisions etc. If an evaluation of a citation index can be carried out in an experimental environment, then the second method of presenting results is probably nearer the real performance of the system, and is used for comparison with the results of the conventional index languages.

As described in some detail in Vol. I (page 110 and Fig. 7.5) the score sheets for each question gave the results with coupling strength from 1 to 7. The basic scoring at the seven coupling levels for the 42 aerodynamics questions with the 1400 document collection is shown in Fig. 7.1T and the results for this set of questions are presented in Fig. 7.2T. Fig. 7.3T presents the results for the 42 questions dealing with structures, while Fig. 7.4T gives the results for the 35 questions having 7 starting terms. A comparative plot of these three question sets is given in Fig. 7.5P.

All the results so far shown are obtained with documents of relevance grades 1-4; Figs. 7.6T, 7.7T and 7.8T show the results for documents of relevance grades 1-3, relevance grades 1-2 and relevance grade 1, with the 42 aerodynamics questions. Fig. 7.9P plots the results of the four grades of relevance.

QUES-TION	COUPLING LEVEL 1		2		3		4		5		6		7	
	R	NR	R	NR	R	NR	R	NR	R	NR	R	NR	R	NR
79	2	136	2	53	0	15	0	3						
100	4	76	4	19	4	10	2	7	2	2	2	0		
116	6	204	6	77	3	33	3	16	3	7	3	5	2	2
118	3	262	3	64	2	23	2	7	0	4				
119	6	317	6	83	4	31	3	10	0	5				
121	-	-	-	-	-	-	-	-	-	-	-	-	-	-
122	2	201	0	45	-	18	-	7	-	5	-	0	-	1
123	-	73	-	15	-	8	-	3	-	3				
126	2	65	2	22	-	8	-	6	-	3	-	1		
130	4	136	2	23	2	10	2	3	-	2	-	1	-	1
132	2	397	-	108	-	40	-	20	-	10	-	6	-	4
136	6	48	5	8	3	5	2	3	2	3	-	2	-	1
137	6	125	4	33	4	11	3	5	3	2	3	-	2	-
141	-	6	-	2	-	2								
145	12	314	12	76	10	38	9	19	9	9	5	4	5	3
146	9	363	9	82	8	45	7	19	7	8	4	4	4	3
147	5	24	3	2	3	-	3	-	2	-	2	-		
148	-	40	-	11	-	3	-	1						
167	4	145	2	36	-	25	-	12	-	3				
170	-	79	-	17	-	3								
181	-	103	-	26	-	12	-	4	-	1	-	1		
182	3	47	2	9	-	4								
189	-	19	-	3	-	1								
190	5	42	5	6	4	1								
223	2	63	2	13	-	1								
224	3	221	2	56	-	8	-	1						
225	4	636	4	199	3	95	3	56	3	39	3	30	3	17
226	7	74	7	15	5	3	3	0	3	0				
227	2	8	2	1	2	1								
230	2	274	-	65	-	27	-	3						
250	7	213	7	62	6	37	5	20	2	11	2	5	2	2
261	-	30	-	8	-	2	-	1	-	1				
264	2	9	2	1	2	1	2	1	2	-	2	-	2	-
266	4	141	3	35	-	14	-	3	-	1	-	1		
268	-	2	-	-										
269	-	-	-	-										
272	3	94	3	23	2	10	2	3	-	2	-	1	-	1
273	6	43	6	7	2	4	2	3	-	3	-	2	-	2
274	-	49	-	12	-	1								
317	2	13	2	4	2	-								
323	2	66	2	13	2	8	-	2	-	1	-	1	-	1
360	4	337	4	112	3	34	3	19	3	10	3	5	3	-

FIGURE 7.1T RESULTS FOR 42 SEARCHES WITH 1400 DOCUMENTS BY BIBLIOGRAPHIC COUPLING

FIGURE 7.2T

Index Language Citation Indexing and Bibliographic Coupling
Document Relevance 1-4

Number of Documents in Collection 1400
Number of Questions 42 (Subset 2 Aerodynamics)
Number of Relevant Documents 198
Generality Number 3.4

Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
1	131	5495*	66.1%	2.3%*	9.345%*
2	113	1446	56.3%	7.2%	2.459%
3	76	592	38.2%	11.4%	1.007%
4	56	257	28.1%	17.9%	0.437%
5	41	135	20.6%	23.3%	0.230%
6	29	70	14.6%	29.3%	0.119%
7+	23	38	11.6%	37.7%	0.065%

FIGURE 7.3T

Index Language Citation Indexing and Bibliographic Coupling
Document Relevance 1-4

Number of Documents in Collection 1400
Number of Questions 42 (Structures)
Number of Relevant Questions 252
Generality Number 4.3

Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
1	179	5962*	71.0%	2.9%*	10.684%*
2	138	1569	54.8%	8.1%	2.668%
3	90	541	35.7%	14.3%	0.920%
4	51	140	20.2%	26.7%	0.238%
5	36	93	14.3%	27.9%	0.193%
6	31	54	12.3%	36.5%	0.092%
7+	25	28	9.9%	47.2%	0.047%

FIGURE 7.4T

Index Language Citation Indexing and Bibliographic Coupling
Document Relevance 1-4

Number of Documents in Collection 1400
Number of Questions 35 (Subset 1)
Number of Relevant Documents 287
Generality Number 5.9

Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
1	195	11658	67.9%*	1.6%*	23.792%*
2	156	3068	54.3%	4.8%	6.261%
3	115	1152	40.1%	9.1%	2.351%
4	79	464	27.5%	14.5%	0.947%
5	56	299	19.5%	15.7%	0.610%
6	43	174	15.0%	19.8%	0.355%
7+	40	101	13.9%	28.4%	0.206%

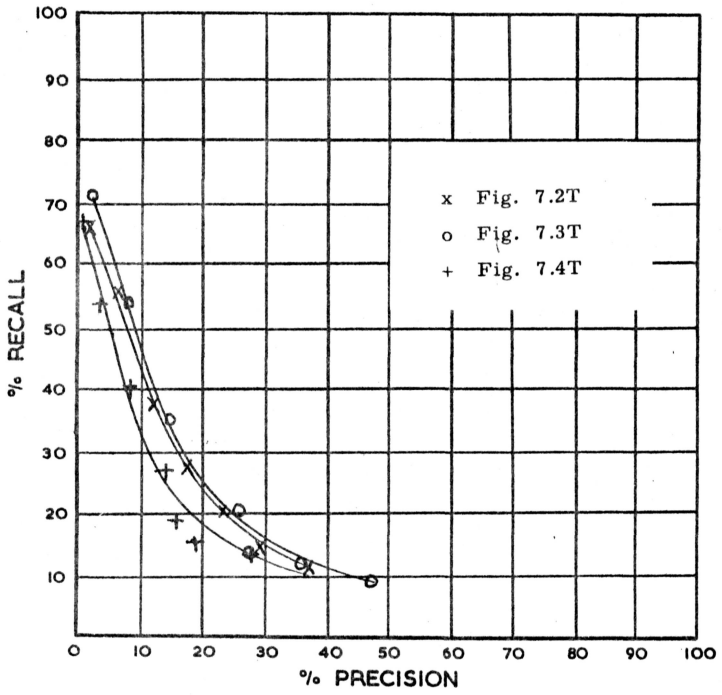

FIGURE 7.5P PLOT OF PERFORMANCE WITH BIBLIOGRAPHIC COUPLING FOR SETS OF 42 AERODYNAMIC QUESTIONS, 42 STRUCTURES QUESTIONS and 35 SEVEN-STARTING-TERM QUESTIONS.

FIGURE 7.6T

Index Language Citation Indexing and Bibliographic Coupling
Document Relevance 1-3

Number of Documents in Collection 1400
Number of Questions 35 (Subset 1)
Number of Relevant Documents 212
Generality Number 4.3

Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
1	139	-	65.5%	-	-
2	106	2021	50.0%	5.0%	4.124%
3	75	759	35.4%	9.0%	1.549%
4	48	370	22.6%	11.5%	0.755%
5	38	187	17.9%	16.9%	0.382%
6	29	102	13.7%	22.1%	0.208%
7+	28	58	13.2%	32.6%	0.118%

FIGURE 7.7T

Index Language Citation Indexing and Bibliograph
Document Relevance 1-2

Number of Documents in Collection 1400
Number of Questions 35
Number of Relevant Documents 79
Generality Number 1.6

Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
1	51	-	64.5%	-	-
2	35	843	44.3%	4.0%	1.720%
3	26	412	32.9%	5.9%	0.841%
4	12	197	15.2%	5.7%	0.402%
5	11	105	13.9%	9.5%	0.214%
6	9	66	11.4%	12.0%	0.135%
7+	8	41	10.1%	16.3%	0.084%

FIGURE 7.8T

Index Language Citation Indexing and Bibliographic Coupling
Document Relevance 1

Number of Documents in Collection 1400
Number of Questions 35
Number of Relevant Documents 18
Generality Number 0.4

Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
1	10	-	55.5%	-	-
2	7	194	38.9%	3.5%	0.396%
3	5	84	27.8%	5.6%	0.171%
4	2	41	11.1%	4.7%	0.084%
5	2	28	11.1%	6.7%	0.057%
6	2	18	11.1%	10.0%	0.037%
7+	2	12	11.1%	14.3%	0.024%

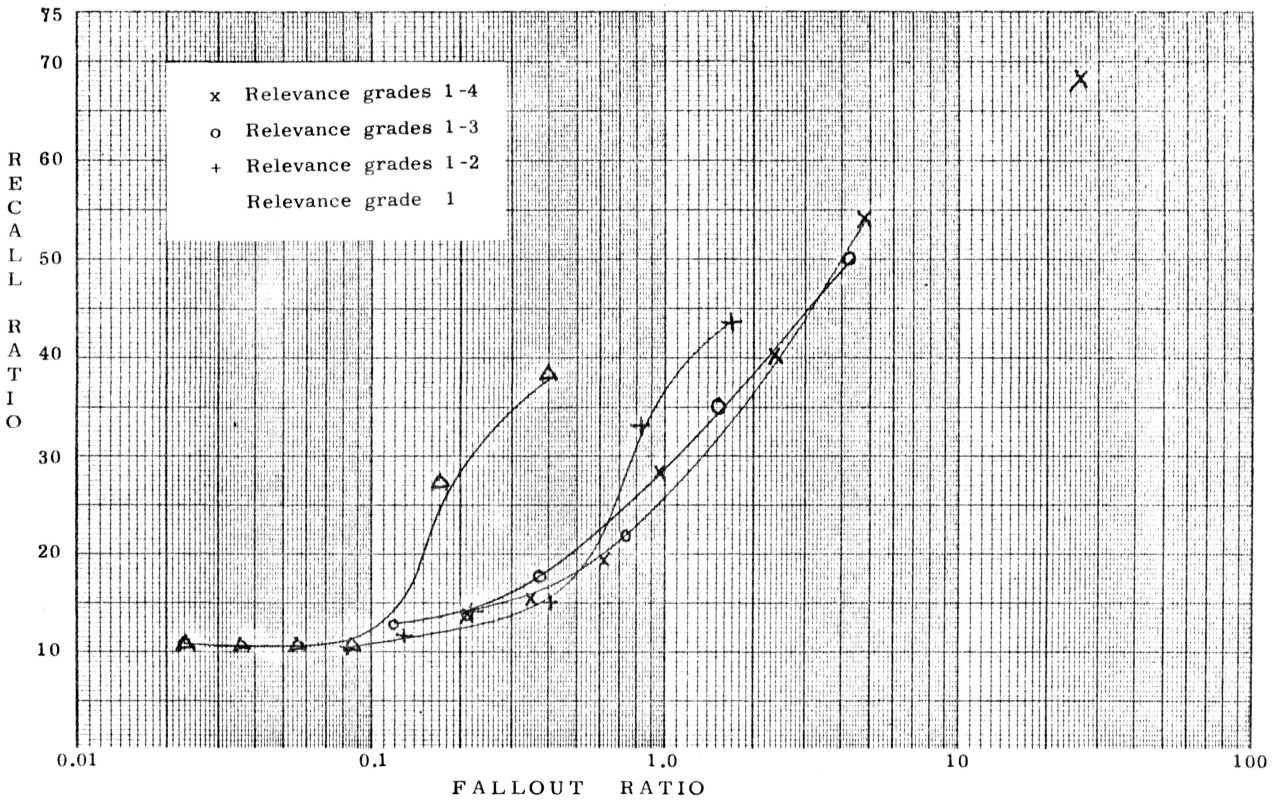

FIGURE 7.9P RECALL/FALLOUT GRAPH FOR PERFORMANCE WITH BIBLIOGRAPHIC COUPLING FOR DOCUMENTS OF FOUR GRADES OF RELEVANCE, WITH 1400 DOCUMENT COLLECTION AND 35 SEVEN-STARTING-TERM QUESTIONS.

In the method used for compiling the scores in the above results, what might be described as the "entry document" was, in the scoring, also counted as a successfully retrieved document; in other words, a previously known relevant document was scored as being a successfully retrieved relevant document. To put it at its simplest, Q227 (as can be seen from Fig. 7.1T) has two relevant documents, the numbers of which were 2087 and 2088. When document 2087 was used as an "entry document", it was found that it had three references in common with document 2088, and therefore both documents were entered as being retrieved at a coupling strength of 3. To take another example Q100 has four relevant documents, numbers 1785, 1786, 1787 and 1788. In the test search, documents 1787 and 1788 were found to have a coupling strength of 6, and documents 1785 and 1786 had a coupling strength of 3. However, there were no references that were common to the pair of documents 1785 and 1786 on the one hand or the pair of documents 1787 and 1788 on the other hand. In spite of this, it would be scored as all four relevant documents having been retrieved at a coupling strength of 3 and lower. As a third example, for Q116 there were six relevant documents, numbers 1317, 1574, 1575, 1576, 1578 and 1656. In the search, document 1576 had a coupling strength of 6 with documents 1574 and 1578, and a coupling strength of 2 with documents 1575 and 1317. In addition document 1317 had a coupling strength of 2 with document 1656. Therefore, at this coupling level this would be recorded as a successful retrieval of all six relevant documents.

By the second method of presenting the results, allowance would be made for these various situations. With Q227, the "entry document" would be eliminated from the scoring; it would be considered that there was only one relevant document, and that this was retrieved. With Q100, however, the first "entry document" would be eliminated from the scoring, but since there was no link between the two pairs of documents, it would be considered that of the three remaining relevant documents, two had been retrieved. With Q116, the "entry document" would be eliminated from the scoring, but since the other five documents were linked either directly or indirectly with the "entry document", all these five documents would be included in the scoring.

On the other hand, with those questions such as Q.122 or Q.132, where no relevant documents were retrieved, the total of relevant documents would in each case be reduced by one.

The result of this exercise is to produce a new set of performance figures where there are now only 156 relevant documents, and the results are presented in Fig. 7.10T. In doing this, it is only the recall and precision ratios that are changed, for the fallout ratio remains the same as in Fig. 7.2T.

It was earlier suggested that it would be reasonable to compare the results by this method with those obtained by the coordination level cut-off. However, as the generality number has been changed, by eliminating 42 relevant documents, it is necessary for this to be done on a recall/fallout graph as in Fig. 7.11P where comparison is made with the Single Term index languages which gave the best and worst performance.

Further tests were carried out where account was taken of the proportional match between documents, this being based on the number of references in the documents concerned. The procedure for doing this was described on pages 110 and 112 of Vol. I. It can make no difference to the

FIGURE 7.10T

Index Language Citation Indexing and Bibliographic Coupling
Document Relevance 1-4

Number of Documents in Collection 1400
Number of Questions 42 (Subset 2)
Number of Relevant Documents 156
Generality Number 2.7

Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
1	95	5495*	60.8%	1.7%*	9.345%*
2	79	1446	50.5%	5.1%	2.459%
3	51	592	32.6%	7.9%	1.007%
4	36	257	23.0%	12.2%	0.437%
5	25	135	16.0%	15.6%	0.230%
6	18	70	11.5%	20.4%	0.119%
7+	12	38	7.6%	24.0%	0.065%

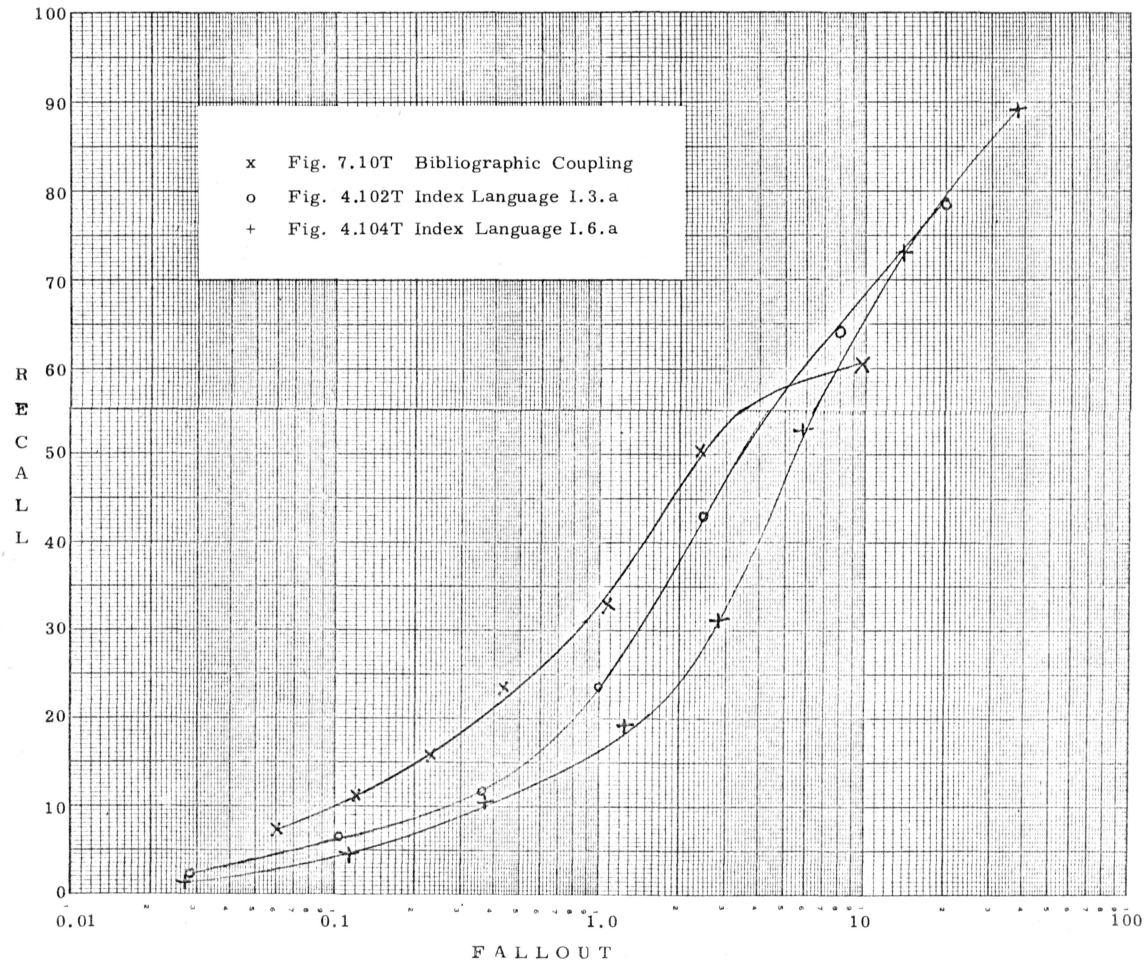

FIGURE 7.11P RECALL FALLOT PLOT FOR BIBLIOGRAPHICAL COUPLING AND SINGLE TERM INDEX LANGUAGES I.3.a AND I.6.a WITH 1400 DOCUMENTS AND 42 AERODYNAMIC QUESTIONS

FIGURE 7.12T

Index Language Citation Indexing and Bibliographic Coupling (Weighted)
Documents Relevance 1-4

Number of Documents in Collection 1400
Number of Questions 42 (Subset 2)
Number of Relevant Documents 198
Generality Number 3.4

Weighted Coupling Strength	Documents Retrieved Rel.	Non-rel.	Recall Ratio	Precision Ratio	Fallout Ratio
150+	131	5495*	66.1%	2.3%*	9.377%*
81-150	122	1758	61.6%	6.5%	3.000%
51-80	111	1387	56.0%	7.4%	2.367%
31-50	90	854	45.5%	9.5%	1.457%
21-30	67	432	33.8%	13.4%	0.737%
16-20	51	207	25.7%	19.7%	0.353%
11-15	38	126	19.1%	23.1%	0.221%
6-10	23	59	11.6%	28.0%	0.101%
3-5	12	20	6.0%	37.5%	0.034%
1-2	2	1	1.0%	66.7%	0.002%

final figure of relevant and non-relevant documents retrieved, but replaces the groups formed at the various coupling levels as given in earlier totals with new groups based on the weighted scores. The results on the 42 aerodynamic questions are shown in Fig. 7.12T; although different groups are formed, there appears to be little variation from the performance for the same document/question set presented in Fig. 7.2T.

As stated in the opening chapter of this volume, we have considerable reservations in presenting these results, in particular when it comes to attempting to make comparison with the performance obtained by conventional methods. One thing that can be stated positively is that the same inverse relationship exists; bibliographic coupling is a precision device which has very much the same effect as coordination in a conventional system.

Since approximately 12% of the documents did not contain any references, it was inevitable that the maximum recall ratio should fall well short of 100%. In the event it appears that, with this collection, something around 70% recall might be expected; for any recall ratio lower than this, the performance appears to compare quite favourably with conventional indexing.