

CHAPTER 6

Supplementary tests and results

Any social agency has a duty to study and evaluate its effectiveness and to seek continuously to improve the methods it employs to achieve its objectives. It is not enough to believe, however sincerely, that we are doing good. It is not enough to invoke 'experience' or to collect meaningless and misleading information... It is not enough to rely upon the support of colleagues and those in the same professional group and to accept their endorsement of our work as proof of its effectiveness. Professional in-group support does not measure effectiveness and does not absolve us from accountability for our decisions. The effectiveness of social agencies, it is claimed, is a question to be determined empirically by methods which can be repeated and verified by others.

L.T. Wilkins: Social Deviance, pages 5 and 6

Whereas in the preceding chapter, the main test results were considered on the basis of the document output cut-off method, with normalised recall ratios, we now return to the basic method used in Chapter 4, and present a series of mainly disconnected notes on various supplementary matters. In some cases, new data are presented; in other cases data which have already been given in Chapter 4 is brought together in different ways in order to illustrate more effectively certain points.

Comparative Results

It is difficult to make direct comparison between the main index languages, because of the inevitable variations created by different numbers of starting terms. However, Fig. 6.1P shows the performance curves for Single Term Natural Language (I.1.a), Simple Concept Natural Language (II.1.a) and Controlled Term, Basic Terms (III.1.a). These might be considered to be comparable since they are all concerned with the basic terms in the particular vocabulary, but the inability of the Simple Concept Index Language to obtain a higher recall figure than 36.9% is due to the severe restrictions which interfixing imposes. That the Controlled Term Index Language also suffers a drop, as compared to Single Term Index Languages, of 7.6% in maximum recall is for the same reason, but the effect is not so severe in this case, since fewer single terms are interfixed. In general the Single Term Natural Language appears to give the best performance.

More reasonable is to make comparison between the index language which have the highest normalised recall ratios in each of the three main groups. These would appear to be Index Languages I.3.a (Single term. Word forms), II.10.a, (Simple Concept. Second alphabetical collateral selected), and III.2.a, (Controlled term. Narrower terms). The results are given in Fig. 6.2P, and show that the Simple Concept index language has made a large increase in maximum recall, but again the Single Term index language appears to give the best performance over the whole curve, thus bearing out the figures presented in Chapter 5.

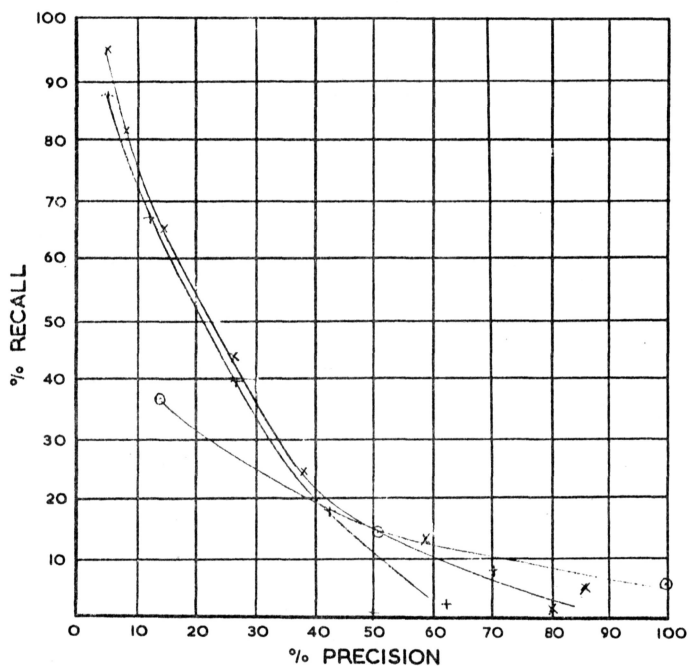


FIGURE 6.1P COMPARISON OF BASIC LANGUAGES
IN THE THREE MAIN GROUPS

x I.1.a. (Fig. 4.140T)
o II.1.a. (Fig. 4.700T)
+ III.1.a. (Fig. 4.800T)

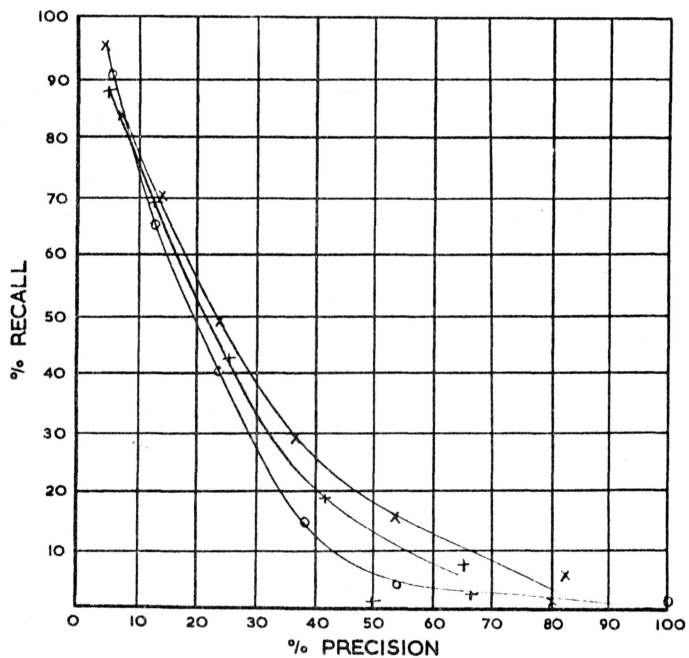


FIGURE 6.2P COMPARISON OF INDEX LANGUAGES
GIVING BEST PERFORMANCE IN
THREE MAIN GROUPS

x I.3.a. (Fig. 4.201T)
o II.10.a. (Fig. 4.709T)
+ III.2.a. (Fig. 4.801T)

Document Relevance

In Chapter 4, Section 6, the effect of relevance was considered, and the results were presented and plotted for documents of different grades of relevance according to the coordination level cut-off. Fig. 6.3T shows the same results as are given in Figs. 4.610T to 4.613T, but now grouped according to relevance grade for each coordination level.

Coordination Level	Relevance Grade	Recall Ratio	Precision Ratio
1	1	94.7	0.3
	1-2	94.2	0.5
	1-3	93.2	0.9
	1-4	94.2	1.1
2	1	85.3	0.7
	1-2	80.6	1.0
	1-3	79.1	2.0
	1-4	77.8	2.4
3	1	60.0	1.2
	1-2	56.1	1.8
	1-3	54.5	3.4
	1-4	48.8	3.6
4	1	42.1	2.2
	1-2	37.4	3.2
	1-3	32.7	5.3
	1-4	29.6	5.8
5	1	25.3	3.1
	1-2	21.3	4.3
	1-3	16.5	6.4
	1-4	16.3	7.6
6	1	14.7	4.5
	1-2	13.5	6.9
	1-3	9.8	9.2
	1-4	9.7	11.4
7	1	7.4	7.1
	1-2	6.5	10.2
	1-3	5.3	16.2
	1-4	5.3	19.2
8	1	3.2	14.3
	1-2	3.9	25.0
	1-3	2.0	25.0
	1-4	1.9	29.2

FIGURE 6.3T RESULTS FOR INDEX LANGUAGE I.1.a FOR 42 QUESTIONS WITH 1400 DOCUMENTS FOR FOUR GRADES OF RELEVANCE.

Plotted as a series of short graphs in Fig. 6.3P, these illustrate yet again the inverse relationship of recall and precision.

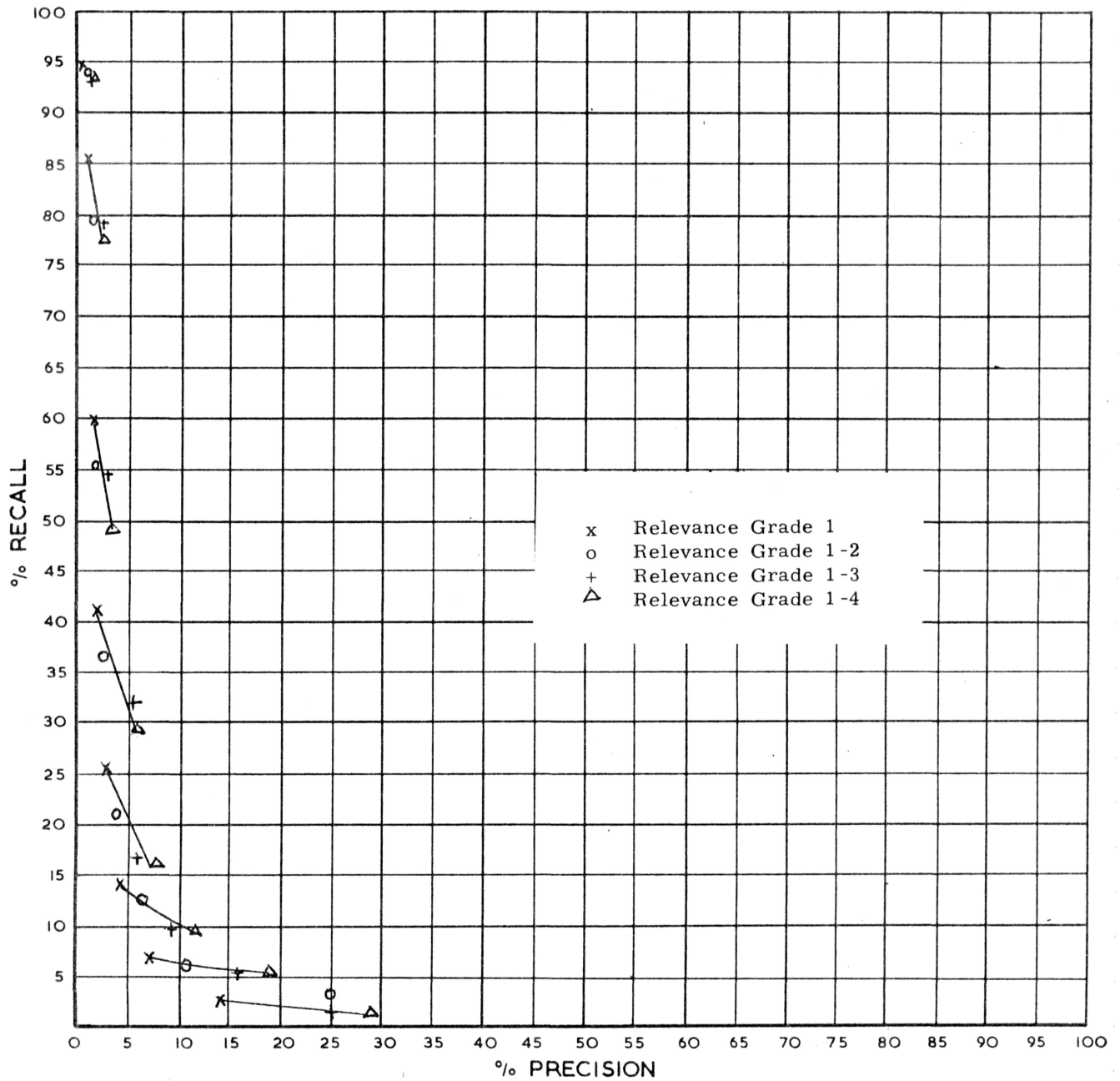


FIGURE 6.3P PLOTS OF EFFECT OF DOCUMENT RELEVANCE
AT EIGHT COORDINATION LEVELS

Basic and supplementary questions

Figs. 6.4T and 6.5T present the results on Index Language I.1.a for the 221 questions when these are divided into the 94 basic questions and 127 supplementary questions (see Vol. 1, Appendix 3G). The basic questions have a generally superior performance, particularly in the middle range of coordination levels and this can be partly accounted for by the higher generality number for this group. On the other hand, documents relevant to the supplementary questions have an average relevance grading that is higher than that for the basic questions (2.7 as against 3.0), and this would have been expected to more than counter the previous effect. It might be suspected that the difference in performance is due to a stronger artificial match between the basic questions and, say, the document titles than exists with the supplementary questions. While analysis does not bear this out, no other adequate explanation can be offered, and the matter is considered again in Chapter 8.

Average of ratios

On pages 51 to 56, the matter of averaging sets of results was considered, the discussion being on the question of using the average of ratios or the average of numbers. To go into this in more detail, the subset of 35 seven-starting-term questions with Index Language I.1.a on the 1400 document collection is used to demonstrate some difficulties that arise with the average of ratios. Numerical results for the 35 questions can be found in Appendix 4A and the results are presented (by the average of numbers) in Fig. 4.110T.

It can be seen from Fig. 6.6T that, when ratios are obtained for each individual question, three different situations arise. Firstly, there are those questions (e.g. Q82) where it is possible to include recall and precision ratios at all coordination levels to the maximum of 7 (since these are all seven-starting-term questions). Secondly, there are those questions (e.g. Q294) where no documents are retrieved at the higher coordination levels, so no ratios can be included. Thirdly, there are those questions (e.g. Q40) where at the higher coordination levels no relevant documents are retrieved although some non-relevant documents are retrieved. This latter situation is indicated in Fig. 6.6T by an asterisk in the appropriate column. Because of these three different situations, it is a matter for argument as to the figure which should be used for obtaining the average ratios. As an example, at the coordination level of four, the sum of the precision ratios is 561.7. In order to obtain the average precision ratio for the whole set of questions, this figure could be divided by 35, this representing the total number of questions. Alternatively it could be divided by 28, representing the questions which, at this particular coordination level, retrieved some documents, either relevant or non-relevant. Finally it could be divided by 23, representing the number of questions which, at this particular coordination level, retrieved relevant documents. With the results by the average of numbers for comparison, the precision ratios obtained by these three methods are given in Fig. 6.7T.

The first method is clearly unsatisfactory; it would appear to be relatively immaterial as to whether method 2 or 3 should be used, but it is obviously important that when results are presented by the average of ratios, it should be made quite clear as to which procedure has been adopted. The complexity involved in presenting results by the average of ratios is an additional reason why, in this report, we have preferred to

FIGURE 6.4T

Index Language I.1.a

Exhaustivity of Indexing 3

Search Rule A

Document Relevance 1-4

Number of Documents in Collection 1400

Number of Questions 94 Basic Questions

Number of Relevant Documents 737

Generality Figure 5.6

Coord- ination Level	Documents Retrieved		Recall Ratio a/a+c	Precision Ratio a/a+b	Fallout Ratio b/b+d
	Rel.	Non-rel.			
1	694	69037	94.2%	1.0%	51.946%
2	604	26777	82.0%	2.2%	20.148%
3	474	10425	64.3%	4.3%	7.844%
4	324	3874	44.0%	7.7%	2.914%
5	179	1286	24.3%	12.2%	0.967%
6	92	333	12.5%	21.6%	0.251%
7	49	112	6.6%	30.8%	0.082%
8	14	24	1.9%	36.8%	0.018%
9	3	3	0.4%	50.0%	0.002%

FIGURE 6.5T

Index Language I.1.a

Exhaustivity of Indexing 3

Search Rule A

Document Relevance 1-4

Number of Documents in Collection 1400

Number of Questions 127 Supplementary Questions

Number of Relevant Documents 853

Generality Figure 4.7

Coord- ination Level	Documents Retrieved		Recall Ratio a/a+c	Precision Ratio a/a+b	Fallout Ratio b/b+d
	Rel.	Non-rel.			
1	816	90085	95.6%	0.9%	48.174%
2	679	31345	79.6%	2.1%	16.762%
3	472	11508	55.3%	3.9%	6.154%
4	282	3485	33.1%	7.5%	1.864%
5	135	1094	15.6%	10.8%	0.585%
6	62	366	7.3%	14.5%	0.195%
7	25	94	2.9%	20.7%	0.049%
8	8	19	0.8%	28.8%	0.010%
9	5	2	0.6%	71.4%	0.001%
10	1	0	0.1%	100.0%	0.000%

QUESTION NUMBER	COORDINATION LEVELS													
	1		2		3		4		5		6		7	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P
2	66.7	3.3	37.5	7.6	12.5	12.0	8.3	100	4.2	10.0				
9	100	0.8	75.0	6.4	25.0	20.0								
34	88.9	1.4	11.1	0.7	-	*	-	*						
40	100	0.2	100	0.5	-	*	-	*	-	*				
49	100	1.6	100	13.9	66.7	40.0								
67	100	1.7	60.0	2.4	30.0	4.1	-	*						
81	100	1.4	93.3	2.4	86.7	3.7	73.3	7.8	40.0	15.0	20.0	42.9		
82	100	2.0	100	4.2	92.9	8.0	85.7	26.7	64.3	60	42.9	66.7	7.1	100
87	100	0.3	100	0.7	100	1.1	100	8.0	25	33.3				
95	100	1.4	100	2.6	70.0	4.2	50	38.5						
113	100	1.4	88.2	3.0	47.1	5.4	23.5	13.8	5.9	100	5.9	100		
122	100	0.6	100	1.5	60.0	3.6	20	11.1						
131	100	1.3	90.9	3.1	54.5	7.9	27.3	27.3	9.1	100				
132	50.0	0.4	50.0	0.8	25.0	1.4	25.0	5.3						
142	100	1.9	100	3.8	91.7	6.4	83.3	12.5	75.0	14.3	33.3	13.3	33.3	33.3
145	100	1.1	100	2.7	92.3	6.6	61.5	9.3	53.8	21.9	38.5	62.5		
157	100	2.2	71.4	8.3	57.1	16.7								
160	100	0.7	100	1.4	80.0	6.3	60.0	16.7	40.0	50.0				
165	100	0.5	75.0	1.2	75.0	3.6	50.0	11.8	25.0	25.0				
170	100	0.3	50.0	0.5	50.0	1.4	50.0	6.3	50.0	33.3				
171	100	0.6	100	2.1	100	4.4	66.7	11.8	-	*	-	*		
177	100	1.5	88.9	7.1	77.8	11.3	55.6	41.7	-	*				
181	100	0.2	100	1.0	50.0	3.8								
189	100	0.6	-	*	-	*	-	*						
205	100	0.4	100	1.3	66.7	2.8	33.3	20.0						
211	100	0.8	100	4.1	83.3	11.1	16.7	20.0						
219	100	2.5	81.8	7.6	72.7	17.0	27.3	23.1	-	*				
261	100	0.5	100	2.3	100	5.8	100	22.2	100	80	100	100	75	100
285	93.8	1.3	87.5	3.2	62.5	4.8	50.0	16.0	18.8	27.3				
292	77.8	1.3	33.3	3.6	22.2	20.0	11.1	100	11.1	100				
293	100	1.4	60.0	2.9	40.0	10.0								
294	92.3	10.7	76.9	40.0										
299	75.0	0.8	41.7	3.2	8.3	11.1								
315	100	1.6	85.7	2.4	71.4	11.8	14.3	11.8						
338	75.0	0.8	75.0	3.2	75.0	13.0	-	*						
Totals	3319.5	49.5	2733.2	151.7	1946.4	279.3	1092.9	561.7	522.2	760.1	240.6	385.4	115.4	233.3

FIGURE 6.6T. PERFORMANCE RESULTS FOR 35 SEVEN-STARTING-TERM QUESTIONS WITH INDEX LANGUAGE I.1.a CALCULATED BY THE AVERAGE OF RATIOS.

Coord- ination Level	Average of numbers	Average of Ratios		
		1 (Total divided by 35)	2 (Total divided by figure shown in brackets)	3 (Total divided by figure shown in brackets)
1	1.1%	1.4%	1.4%(35)	1.4%(35)
2	2.7%	4.3%	4.3%(35)	4.4%(34)
3	5.2%	8.0%	8.2%(34)	9.0%(31)
4	13.5%	16.0%	20.1%(28)	24.4%(23)
5	23.8%	21.7%	42.2%(18)	54.3%(14)
6	37.2%	11.0%	55.0%(7)	64.2%(6)
7	50.0%	6.7%	77.8%(3)	77.8%(3)

FIGURE 6.7T. PRECISION RATIOS OBTAINED BY THREE DIFFERENT AVERAGE OF RATIOS PROCEDURES.

use the average of numbers.

Comparison of documents dealing with aerodynamics and structures

The main sets of test results in Chapter 4 were concerned with a subset of 42 questions all of which dealt with aerodynamics rather than structures. For comparison purposes, a set of 42 questions on structures was prepared. Searched on the 1400 document collection, with index language I.1.a, the tests results are given in Fig. 6.8T. Comparison is made in Fig. 6.9P with the results as given in Fig. 4.120T for the 42 aerodynamic questions under the same conditions. This plot shows an unusual characteristic, in that at the higher recall levels, the structure questions have superior precision, but at a recall ratio of about 25%, the curves cross over, and the aerodynamic questions have the better performance.

There are two reasons why one would expect the structure questions to do better. Firstly there are more relevant documents, and therefore the generality number is higher, namely 4.3 as against 3.4. Secondly, although to calculate the generality number N is presumed to be 1400, real N must (as argued on pages 71 - 76) be considerably less than this number.

If the position at a coordination level of 3 is considered, the performance figures are as follows:

Aerodynamics (As Fig. 4.120T)			Structures (As Fig. 6.9T)		
Recall Ratio	Precision Ratio	Fallout Ratio	Recall Ratio	Precision Ratio	Fallout Ratio
66.7%	3.2%	6.790%	67.5%	8.6%	1.732%

To allow for the difference in the generality number, the precision ratio for the aerodynamic questions can be adjusted by the equation given on page 73 and this would result in a new precision ratio of 4.1% which continues to be well below the comparable figure for the structures questions.

FIGURE 6.8T

Index Language I.1.a

Exhaustivity of Indexing 3

Search Rule A

Document Relevance 1-4

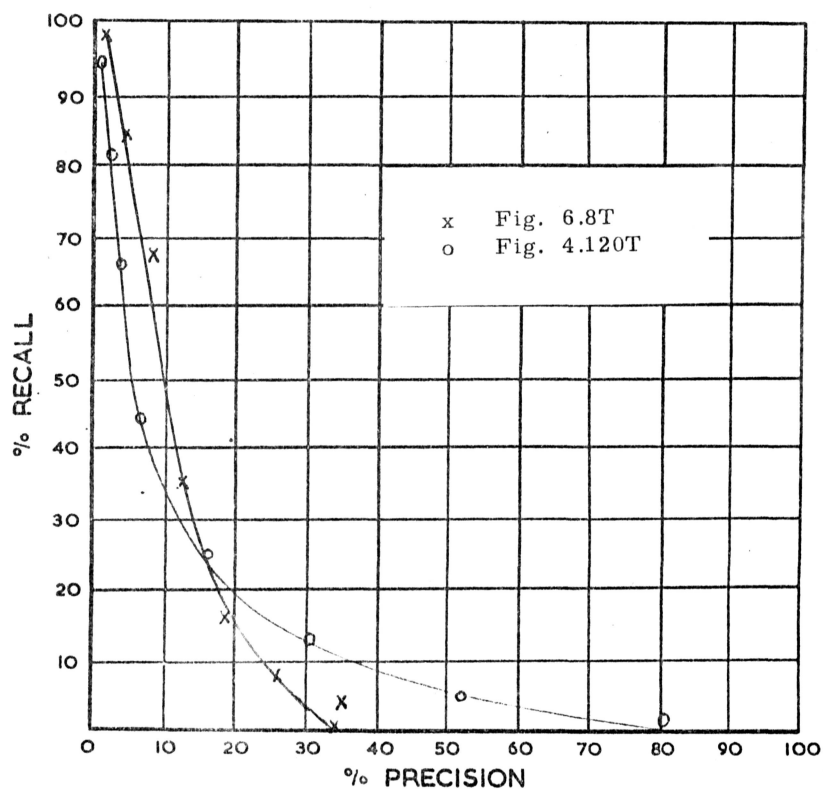
Number of Documents in Collection 1400

Number of Questions 42 Structures Questions

Number of Relevant Documents 255

Generality Number 4.3

Coord- ination Level	Documents Retrieved		Recall Ratio a/a+c	Precision Ratio a/a+b	Fallout Ratio b/b+d	x	y	z
	Rel.	Non-rel.						
1	251	22629	98.4%	1.1%	38.652%	42	42	42
2	216	6081	84.7%	3.4%	10.387%	42	42	42
3	172	1822	67.5%	8.6%	3.112%	41	42	42
4	92	602	36.1%	13.3%	1.028%	34	41	41
5	43	182	16.9%	19.1%	0.311%	23	37	37
6	23	65	9.0%	26.1%	0.111%	15	30	30
7	12	21	4.7%	36.4%	0.036%	10	25	25
8	2	4	0.8%	33.3%	0.007%	2	19	19
9	1	2	0.4%	33.3%	0.003%	2	14	14
10	0	0				0	10	10
11	0	0				0	4	4



Since this direct adjustment on the basis of generality does not equate performance, it is therefore necessary to consider whether N should be revised for the structures questions. It has already been established (on pages 71 to 76) that real N for the aerodynamics questions is not 1400, but in the region of 1027, at which figure the true generality number is 4.6. One might first hypothesise that the remaining 373 documents represent N for the structures questions; the corresponding generality number would be 16.3. To match this new figure, the adjusted precision ratio for the aerodynamics questions would now be 10.6%, which is higher than the figure (8.6%) for the structures questions. It therefore appears that in the collection of 1400 documents, there must be a subset which is common to both. Using the methods given on pages 71 to 76, N for the structures questions is shown to be probably at least 474, which gives a generality number of 12.8. Adjusted to this generality number the precision ratio for the aerodynamics questions is now 8.6%, the same as for the structures questions. The fallout ratios also now match; for the aerodynamics questions, where $N = 1027$, the fallout ratio is 9.2%; for the structures questions, where $N = 474$, the fallout ratio is 9.270%.

The phrase "probably at least 474" was used because no account has been taken of the possibility that the performance figures will be affected by the comparative firmness of the terminology of aerodynamics and structures. The phrase, in fact, implied a belief that aerodynamics has the mushier or more imprecise language, and that for this reason, one would expect the set of structures questions to provide the better performance.

However, the matter is complicated even if this latter point is ignored. At a coordination level of five, the structures questions have a performance of 16.1% recall and 18.1% precision. No exact matching figures can be obtained from Fig. 4.120T, but reference to 4.125P shows that, for the aerodynamic questions, at 16% recall, precision would be approximately 25%. Adjusted for generality on the basis worked out earlier, this would increase the precision ratio to 62%, which is far in advance of the figures for the structures questions. On the other hand at a single term level, it is found that the 42 structures questions have retrieved a total of 22,929 documents, which is an average of 538 documents for each question. This is a figure larger than the 474 documents earlier hypothesised as representing N.

The above discussion is neither clear nor conclusive, and offers no explanation for the crossover in the performance figures of the two sets of questions (which is probably an aberration caused by the relatively small number of results). Rather it serves to point up some of the difficulties which are involved in attempting to compare performance in different subject areas by the coordination level cut-off, and emphasises the necessity for further research in this and related fields.

Performance comparison by coordination levels

In Chapter 4, all the tables of results and accompanying performance curves were based on the variation of coordination level. From these tables, sets of figures are extracted where the coordination level is held constant while the variable is the index language. Figs. 6.10T and 6.11T deal with the Single Term index languages at a coordination level of 3 and 6. Figs. 6.12T and 6.13T present the results at coordination levels of 2 and 4 for the Simple Concept index languages, while Figs. 6.14T and 6.15T present results at the same coordination levels for the Controlled Term index languages.

FIGURE 6.10T

Coordination Level 3

Exhaustivity 3

Document Relevance 1-4

Search Rule A

Number of Documents in Collection 200

Number of Questions 42

Number of Relevant Documents 198

Generality Figure 23.6

Index Language	Documents Retrieved		Recall Ratio	Precision Ratio	Fall-out Ratio	x	y	z
	Rel.	Non-rel.						
I.1.a	132	761	66.7%	14.8%	9.278%	42	42	42
I.2.a	134	837	67.7%	13.8%	10.205%	42	42	42
I.3.a	139	913	70.2%	13.2%	11.131%	42	42	42
I.5.a	144	1613	72.7%	8.2%	19.666%	42	42	42
I.6.a	151	1785	76.3%	7.8%	21.763%	42	42	42
I.7.a	143	989	72.2%	12.6%	12.058%	42	42	42
I.8.a	146	1081	73.7%	11.9%	13.180%	42	42	42
I.9.a	163	2034	82.3%	7.4%	24.799%	42	42	42

FIGURE 6.11T

Coordination Level 6

Exhaustivity 3

Document Relevance 1-4

Search Rule A

Number of Documents in Collection 200

Number of Questions 42

Number of Relevant Documents 198

Generality Figure 23.6

Index Language	Documents Retrieved		Recall Ratio	Precision Ratio	Fall-out Ratio	x	y	z
	Rel.	Non-rel.						
I.1.a	25	17	12.6%	59.5%	0.207%	15	33	33
I.2.a	25	23	12.6%	52.1%	0.280%	15	33	33
I.3.a	32	28	16.2%	53.3%	0.341%	17	33	33
I.5.a	41	69	20.7%	37.3%	0.841%	23	33	33
I.6.a	47	87	23.7%	35.1%	1.061%	23	33	33
I.7.a	31	31	15.7%	50.0%	0.378%	20	33	33
I.8.a	33	37	16.7%	47.1%	0.451%	21	33	33
I.9.a	46	171	23.2%	21.2%	2.085%	25	33	33

FIGURE 6.12T

Coordination Level 2

Exhaustivity 3

Document Relevance 1-4

Search Rule A

Number of Documents in Collection 200

Number of Questions 42

Number of Relevant Documents 198

Generality Figure 23.6

Index Language	Documents Retrieved		Recall Ratio	Precision Ratio	Fall-out Ratio	x	y	z
	Rel.	Non-rel.						
II.1.a	28	27	14.1%	50.9%	0.329%	18	41	41
II.2.a	44	65	22.2%	40.4%	0.792%	25	41	41
II.3.a	60	115	30.3%	27.9%	1.890%	32	41	41
II.4.a	67	255	33.8%	20.8%	3.109%	36	41	41
II.5.a	81	359	40.9%	18.4%	4.377%	39	41	41
II.6.a	57	155	28.8%	26.9%	1.890%	33	41	41
II.7.a	75	318	37.9%	19.1%	3.877%	33	41	41
II.8.a	99	605	50.0%	14.1%	7.376%	40	41	41
II.9.a	75	296	37.9%	20.2%	3.609%	33	41	41
II.10.a	129	942	65.2%	12.0%	11.485%	32	41	41
II.11.a	146	1259	73.7%	10.4%	15.350%	41	41	41
II.12.a	72	287	36.4%	20.1%	3.500%	34	41	41
II.13.a	117	937	59.1%	11.1%	11.424%	40	41	41
II.14.a	143	2047	72.2%	6.5%	24.957%	40	41	41
II.15.a	168	2590	84.8%	6.1%	31.578%	41	41	41

FIGURE 6.13T

Coordination Level 4

Exhaustivity 3

Document Relevance 1-4

Search Rule A

Number of Documents in Collection 200

Number of Questions 42

Number of Relevant Documents 198

Generality Figure 23.6

Index Language	Documents Retrieved		Recall Ratio	Precision Ratio	Fall-out Ratio	x	y	z
	Rel.	Non-rel.						
II.1.a	0	0				0	32	32
II.2.a	2	0	1.0%	100%	0.000%	2	32	32
II.3.a	4	0	2.0%	100%	0.000%	3	32	32
II.4.a	2	8	1.0%	20.0%	0.098%	4	32	32
II.5.a	6	10	3.0%	37.5%	0.122%	5	32	32
II.6.a	4	5	2.0%	44.4%	0.061%	4	32	32
II.7.a	6	9	3.0%	40.0%	0.110%	4	32	32
II.8.a	14	33	7.1%	29.8%	0.402%	10	32	32
II.9.a	13	6	6.6%	68.4%	0.073%	8	32	32
II.10.a	27	44	13.6%	38.0%	0.536%	18	32	32
II.11.a	34	99	17.2%	25.6%	1.207%	19	32	32
II.12.a	9	3	4.5%	75.0%	0.037%	5	32	32
II.13.a	20	54	10.1%	27.0%	0.658%	11	32	32
II.14.a	26	179	13.1%	12.7%	2.182%	19	32	32
II.15.a	51	325	25.8%	13.6%	3.962%	26	32	32

FIGURE 6.14T

Coordination Level 2

Exhaustivity 3

Document Relevance 1-4

Search Rule A

Number of Documents in Collection 200

Number of Questions 42

Number of Relevant Documents 198

Generality Figure 23.6

Index Language	Documents Retrieved		Recall Ratio	Precision Ratio	Fall-out Ratio	x	y	z
	Rel.	Non-rel.						
III.1	136	946	68.7%	12.6%	11.534%	42	42	42
III.2	137	1024	69.2%	11.8%	12.485%	42	42	42
III.3	147	1575	74.2%	8.5%	19.203%	42	42	42
III.4	148	1661	74.7%	8.2%	20.251%	42	42	42
III.5	184	3044	92.9%	5.7%	37.113%	42	42	42
III.6	187	3482	94.4%	5.1%	42.465%	42	42	42

FIGURE 6.15T

Coordination Level 4

Exhaustivity 3

Document Relevance 1-4

Search Rule A

Number of Documents in Collection 200

Number of Questions 42

Number of Relevant Documents 198

Generality Figure 23.6

Index Language	Documents Retrieved		Recall Ratio	Precision Ratio	Fall-out Ratio	x	y	z
	Rel.	Non-rel.						
III.1	36	49	18.2%	42.4%	0.597%	20	34	34
III.2	38	53	19.2%	41.8%	0.646%	20	34	34
III.3	43	78	21.7%	35.5%	0.951%	22	34	34
III.4	44	82	22.2%	34.9%	1.000%	23	34	34
III.5	62	420	31.3%	12.9%	5.121%	19	34	34
III.6	72	527	36.4%	12.0%	6.425%	31	34	34

If one makes the assumption that the coordination level of 3 for the Single Term index languages is approximately equal to a coordination level of 2 for the Simple Concept and Controlled Term index languages, then it is possible to present in a bar chart a representation of what happens in regard to recall and precision ratios when moving from one index language to another. Index Language II.1.a has the lowest recall ratio and highest precision ratio so this is taken as the starting point in Fig. 6.16T.

Effects of precision devices

In Chapter 4, Section 3, the results of tests on the Single Term index languages with interfixing and partitioning were presented. Figure 6.17T and 6.18T make extracts from these tables of the figures at the coordination level of 4.

Effects of question generality

The individual results for each of the 221 questions with the 1400 document collection and Index Language I.1.a are given in Appendix 4A, and the figures for this particular set of results are given in Fig. 4.100T. As discussed in Chapter 3, this set of questions was a heterogeneous group in a number of respects; various breakdowns have now been made.

First the questions have been grouped according to the number of documents relevant to each question, and table 6.19T shows the recall and precision ratios for each of the groups.

There appears to be a general trend towards a lower recall ratio at any given coordination level for those searches where there are increased numbers of relevant documents; as usual this is matched by a higher precision ratio. If the questions having 1-4 relevant documents and the questions having 16 or more relevant documents are grouped, then this change becomes more apparent, as is shown in Fig. 6.20T.

However, the marked increase in the precision ratio at any given recall ratio is obviously due to the large increase in the generality number of the questions having 16 or more relevant documents. If one considers the fallout ratio, it can be seen from Fig. 6.21P that when recall is plotted against fallout, those questions which have four or less relevant documents have markedly superior performance.

It would probably be correct to say that, as a rule, a question having few relevant documents is a specific query, while a question having a very large number of relevant documents is likely to be a general question. From this it is reasonable to hypothesise that a specific question should present a simpler retrieval problem a general question. Without suggesting that the results presented above prove this hypothesis, they can certainly be said to support it.

Effect of number of postings

The next breakdown of the 221 questions was made by grouping the questions according to the numbers of total postings of the question search terms; information on this is included with the set of results in Appendix 4A. For instance, as can be checked from Appendix 5.1 of Volume I, the three search terms of Question 295 (i.e. 'uniformly', 'loaded', 'sectors'.) have a total of only 46 postings, while for Question 106, the nine search terms have a total of 3,474 postings. Ten groups were formed on this basis, each

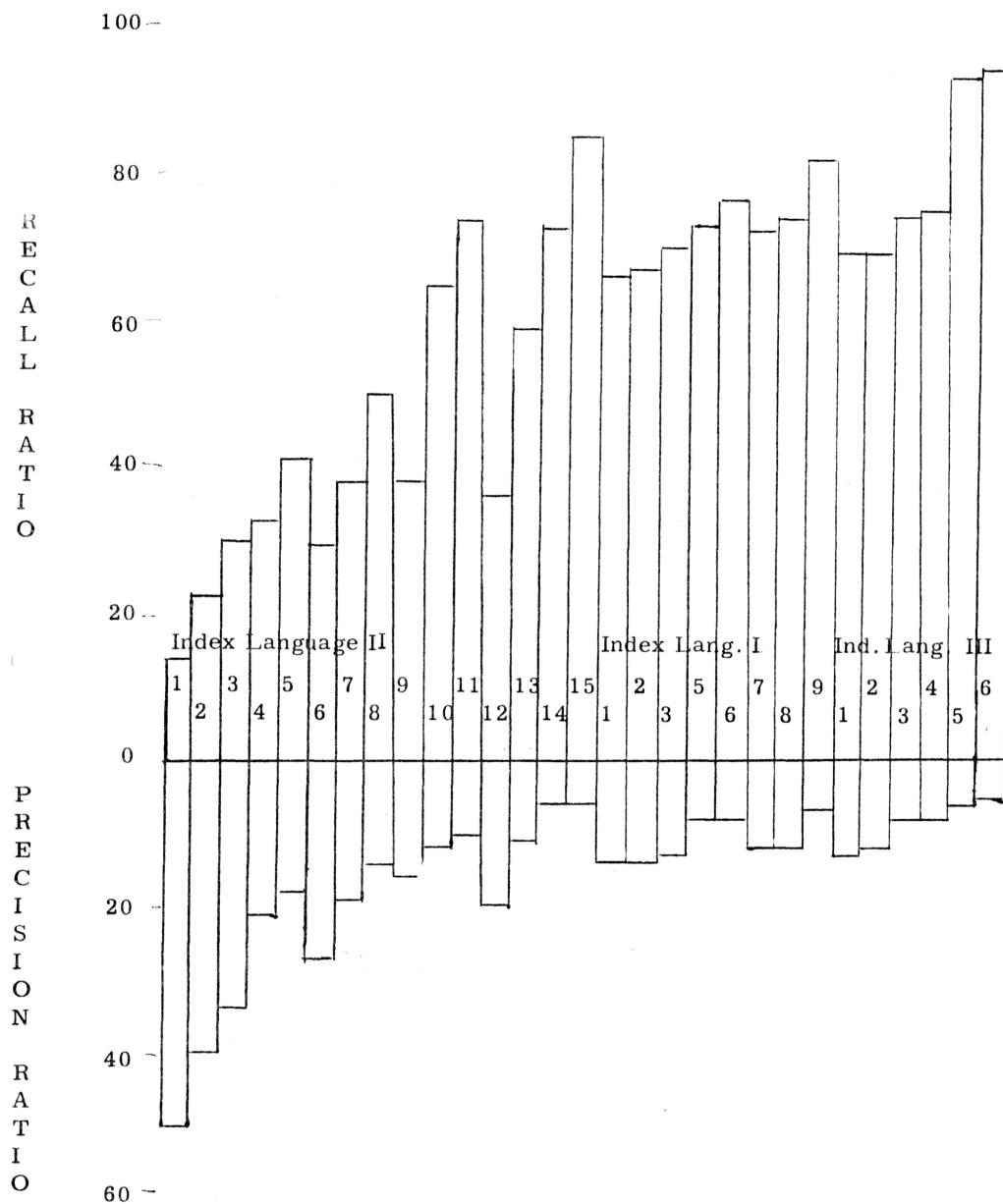


FIGURE 6.16T RECALL AND PRECISION RATIOS WITH 200 DOCUMENTS AND 42 QUESTIONS AT COORDINATION LEVEL OF 2 FOR SIMPLE CONCEPT AND CONTROLLED TERM INDEX LANGUAGES; AND AT COORDINATION LEVEL OF 3 FOR SINGLE TERM INDEX LANGUAGES.

FIGURE 6.17T

Exhaustivity of Indexing 3
Search Rule B
Document Relevance 1-4

Number of Documents 1400
Number of Questions 19 (Subset 4)
Number of Relevant Documents 131
Generality Number 4.9

Coordination Level 4

Figure	Index Language	Documents Retrieved Rel. Non-rel.		Recall Ratio	Precision Ratio	Fallout Ratio	x y z
4.300T	I.1.a	37	127	28.2%	22.6%	0.480%	15 19 19
4.310T	I.1.b (Partitioning)	26	75	19.8%	25.7%	0.283%	12 19 19
4.320T	I.1.c (Interfixing)	25	54	19.1%	31.6%	0.204%	12 19 19
4.330T	I.1.d (Interfixing & Partitioning)	16	36	12.2%	30.8%	0.136%	10 19 19

FIGURE 6.18T

Exhaustivity of Indexing 3
Search Rule B
Document Relevance 1-4

Number of Documents 1400
Number of Questions 19 (Subset 4)
Number of Relevant Documents 131
Generality Number 4.9

Coordination Level 4

Figure	Index Language	Documents Retrieved Rel. Non-rel.		Recall Ratio	Precision Ratio	Fallout Ratio	x y z
4.302T	I.6.a	58	765	44.3%	7.0%	2.890%	19 19 19
4.312T	I.6.b (Partitioning)	39	321	29.8%	10.8%	1.213%	16 19 19
4.322T	I.6.c (Interfixing)	38	312	29.0%	10.9%	1.179%	17 19 19
4.332T	I.6.d (Interfixing & Partitioning)	27	148	20.6%	15.4%	0.559%	12 19 19

No. of relevant documents per question	No. of questions in group	COORDINATION LEVEL									
		1	2	3	4	5	6	7	8	9	10
1	6	100 0.1	83.3 0.2	66.7 0.3	50.0 1.5	29.0 9.0	14.5 18.8	12.9 29.6	3.2 28.6	1.6 33.7	
2	31	95.2 0.3	79.0 0.7	69.4 1.7	45.2 4.0	24.1 7.4	18.5 14.5	9.3 38.5	3.7 100		
3	18	98.1 0.4	94.4 1.3	66.7 2.8	46.3 5.5	23.0 17.0	11.0 36.7	6.0 66.7	2.0 100		
4	25	92.0 0.5	76.0 1.1	58.0 2.5	37.0 6.5	18.5 6.9	7.7 9.9	4.6 25.0	2.3 100	0.8 100	
5	26	100 0.6	90.0 1.5	66.2 3.0	39.2 4.8	41.2 9.1	26.3 15.1	13.2 15.2	5.3 24.0	3.5 57.1	0.9 100
6	20	97.4 0.7	90.4 1.7	78.9 3.5	62.3 6.0	12.5 18.2	3.6 17.4	2.7 50.0	0.9 50.0		
7	16	98.2 1.2	76.8 3.4	52.7 6.2	25.0 9.9	16.1 7.4	6.3 7.4	0.9 11.1			
8	14	97.3 1.2	85.7 2.7	59.8 3.9	35.7 6.6	16.3 19.5	8.1 36.7	5.2 53.8	0.7 50.0	0.7 100	
9	15	97.8 1.3	83.7 3.7	59.3 8.6	35.6 14.9	11.4 14.5	8.6 25.0	2.9 50.0	2.9 66.7	1.4 100	
10	7	94.3 1.2	84.3 2.3	57.1 4.0	35.7 9.5	10.4 22.1	10.4 22.1	3.5 21.9	0.5 11.8		
11-15	30	93.3 3.2	82.6 3.7	61.0 6.0	38.2 9.7	20.1 13.4	4.3 30.0	2.1 100			
16-20	8	99.3 1.6	87.2 3.1	61.7 6.2	44.0 11.7	18.4 19.7	4.2 15.4				
21-25	2	81.3 3.1	45.8 4.3	16.7 4.5	10.4 10.9	7.1 100	3.6 100				
26-30	1	78.6 5.5	50.0 9.3	32.1 31.0	10.7 100	15.3 25.6	9.7 38.9	4.2 75.0			
31-40	2	75.0 4.5	51.4 5.6	40.3 9.2	33.3 14.2						

FIGURE 6.19T RESULTS OF 221 SEARCHES ON 1400 DOCUMENTS WITH INDEX LANGUAGE I.1.a GROUPED ACCORDING TO NUMBER OF DOCUMENTS RELEVANT TO QUESTION.

1-4 RELEVANT DOCUMENTS				16 OR MORE RELEVANT DOCUMENTS		
COORD- INATION LEVEL	RECALL RATIO	PRECISION RATIO	FALLOUT RATIO	RECALL RATIO	PRECISION RATIO	FALLOUT RATIO
1	96.7%	0.3%	50.181%	77.7%	6.2%	19.670%
2	86.1%	0.9%	15.635%	49.3%	5.5%	14.042%
3	68.1%	1.7%	6.382%	31.1%	8.8%	5.378%
4	45.9%	4.1%	1.687%	21.6%	14.7%	2.101%
5	25.4%	8.1%	0.473%	10.1%	25.9%	0.486%
6	15.6%	16.1%	0.128%	5.4%	38.1%	0.147%
7	10.7%	32.5%	0.035%	2.0%	75.0%	0.010%
8	1.6%	28.6%	0.006%			
9	0.8%	33.3%	0.002%			

FIGURE 6.20T RESULTS AS FIGURE 6.19T GROUPED
FOR QUESTIONS WITH FEW AND MANY
RELEVANT DOCUMENTS

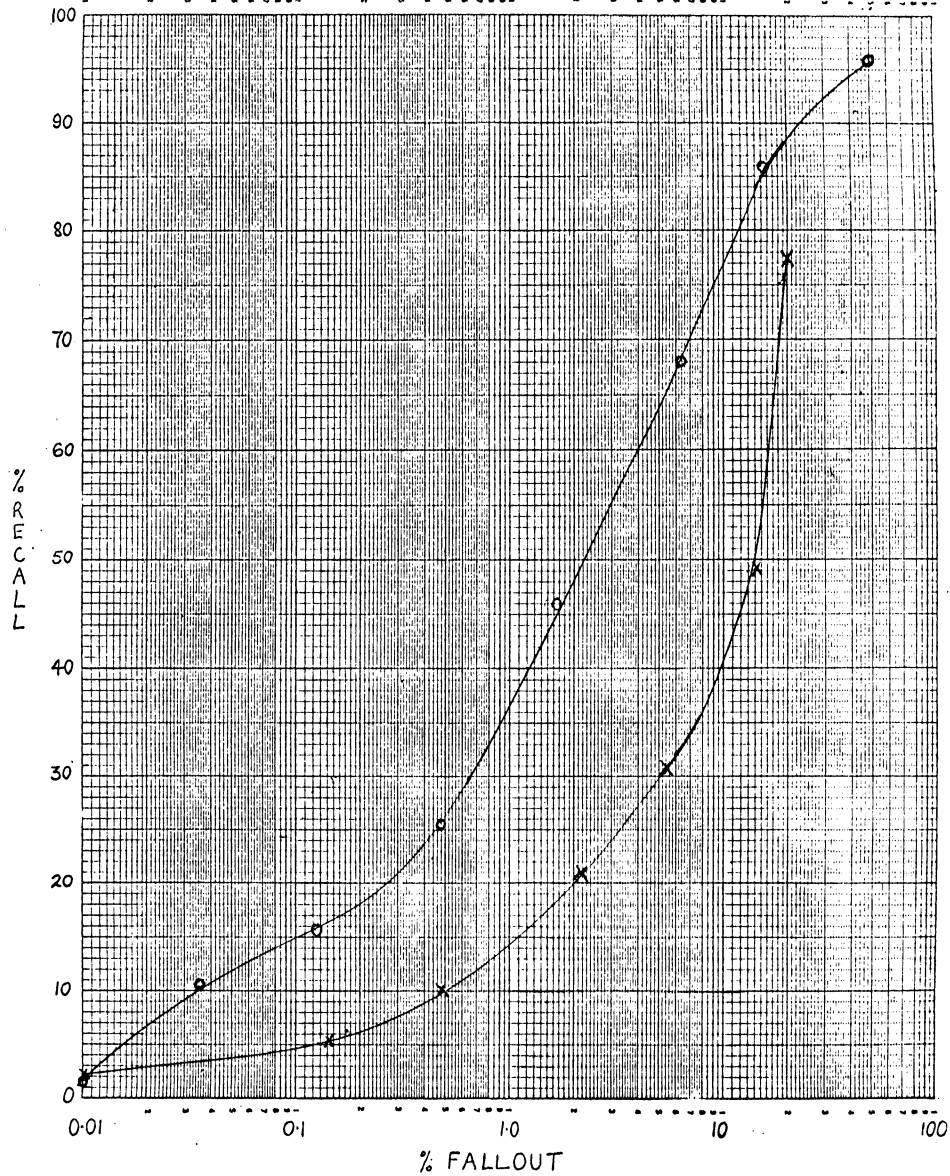


FIGURE 6.21P RECALL/FALLOUT PLOT OF RESULTS
PRESENTED IN FIGURE 6.20T

containing approximately similar numbers of questions; the results are presented in Fig. 6.22T.

With some minor aberrations, these show, for any given coordination level, that the increase in total term postings results in a regular increase in recall ratio accompanied by a corresponding decrease in precision ratio. The next stage was to group the questions in relation to the average number of postings for each term. However, the preliminary stage of making up the groups of questions by this method showed that the groups differed little from those used in Fig. 6.22T, so no further work on this was done.

Order of retrieval of relevant documents

An analysis was made of the effect on retrieval of individual relevant documents in moving from one index language to another. The results in Chapter 4 show, for instance, that with Index Language I.1.a, at a coordination level of 5, there were 94 relevant documents retrieved (see Fig. 4.200T). With Index Language I.6.a, at a coordination level of 6, there were 87 relevant documents retrieved (see Fig. 4.203T). The question is whether the change of index language and the increase in coordination levels resulted in a different or a similar set of retrieved documents.

To investigate this point, nine index languages were selected namely I.1.a, I.5.a, I.8.a, II.1.a, II.5.a, II.10.a, II.15.a, III.1.a and III.6.a. For the 42 questions, the records were checked to find the order of retrieval of the relevant documents for each of the nine languages. Some examples are given in Fig. 6.23T, which shows, for Questions 118, 170 and 250, the coordination levels at which the relevant documents were retrieved, and in Fig. 6.24T a ranked order of retrieval. From this type of data for the 42 questions, it was too involved to sort out what happened to each individual relevant document, but an analysis was made for each of the three main groups of index languages to find what happened to the relevant documents ranked first and last in the basic languages (i.e. I.1.a, II.1.a, and III.1.a). While it was not possible to make a clear cut decision every time, Fig. 6.25T shows that in the very large majority of cases, the change from one language to another did not alter the retrieval rank of the first and last documents retrieved.

QUESTION NUMBER	INDEX LANGUAGE	COORDINATION LEVEL							
		1	2	3	4	5	6	7	8
118	I.1			1324 1378		1667 1666 1670			
	I.5			1378		1324	1667 1670	1666	
	I.8			1324	1378	1666	1667 1670		
	II.1	1667 1324 1378		1666 1670					
	II.5	1667 1378	1324		1666 1670				
	II.10		1324 1378		1666 1670	1667			
	II.15			1324 1378	1666 1670	1667			
	III.1		1324 1378		1667 1666 1670				
	III.6			1324 1378	1667 1666 1670				
170	I.1	1605				1360			
	I.5			1605			1360		
	I.8		1605		1360				
	II.1	1360							
	II.5	1605			1360				
	II.10	1605			1360				
	II.15		1605		1360				
	III.1		1605		1360				
	III.6		1605		1360				
250	I.1				2364 2367 1335	1798 1316 1415 1416	1311		
	I.5				2367 1335	1798 2364 1316 1416	1311 1415		
	I.8				2367 1335	1798 2364 1316 1416	1311 1415		
	II.1	1415 2364 1335	1311 1798 1316						
	II.5		1416 1798 2364 2367 1335	1311 1415 1316					
	II.10	2367	1798 2364 1316 1335	1311 1415 1416					
	II.15		1798 2364 2367 1335	1311 1415 1416 1316					
	III.1		1335 2367	2364	1316 1798	1311 1415 1416			
	III.6		1335	2364 2367	1316 1798	1311 1415 1416			

FIGURE 6.23T COORDINATION LEVEL OF RETRIEVAL OF
RELEVANT DOCUMENTS FOR QUESTIONS
118, 170 AND 250 BY NINE INDEX LANGUAGES

QUESTION NUMBER	RELEVANT DOCUMENTS	INDEX LANGUAGE								
		I.1	I.5	I.8	II.1	II.5	II.10	II.15	III.1	III.6
118	1324	4=	4	5	3=	3	4=	4=	4=	4=
	1378	4=	5	4	3=	4	4=	4=	4=	4=
	1666	1=	1	3	1=	1=	2=	2=	1=	1=
	1667	1=	2=	1=	3=	4=	1	1	1=	1=
	1670	1=	2=	1=	1=	1=	2=	2=	1=	1=
170	1360	1	1	1	1	1	1	1	1	1
	1605	2	2	2	2	2	2	2	2	2
250	1311	1=	1=	1=	1=	1=	1=	1=	1=	1=
	1316	4=	4=	4=	1=	1=	4=	1=	4=	4=
	1335	6=	7=	7=	4=	4=	4=	5=	7=	8=
	1415	1=	1=	1=	4=	1=	1=	1=	1=	1=
	1416	1=	1=	1=	7=	4=	1=	1=	1=	1=
	1798	4=	4=	4=	1=	4=	4=	5=	4=	4=
	2364	6=	4=	4=	4=	4=	4=	5=	6	6=
	2367	6=	7=	7=	7=	4=	8	5=	7=	6=

FIGURE 6.24T RANKED ORDER OUTPUT FOR RELEVANT DOCUMENTS OF QUESTIONS 118, 170 AND 250 BY NINE INDEX LANGUAGES

INDEX LANGUAGE GROUP		HIGHEST RANKED DOCUMENTS		LOWEST RANKED DOCUMENTS	
		Maintained Position	Changed Position	Maintained Position	Changed Position
I	SINGLE TERMS	33	3	33	1
II	SIMPLE CONCEPTS	26	7	28	2
III	CONTROLLED TERMS	30	6	29	4

FIGURE 6.25T EFFECT ON RANK OF HIGHEST AND LOWEST RANKED DOCUMENTS IN MOVING TO DIFFERENT INDEX LANGUAGE FOR 42 QUESTIONS