CHAPTER 5

Simulated ranking and document output cut-off

> There is confusion of ends and means in this type of attack upon measurement in principle. Perhaps if medicine threw away the thermometer, the encephalograph, the X-ray, and all other technicalities, medicine would become much more human! How much more preferable the tender hand on the brow than a nasty piece of glass in the mouth - how inhuman! But is it sympathy and fellow-feeling that we want from the physician or a technical competence to identify the condition and give us the cure? The bedside manner still has a place in the cure, even although the hand on the brow has been replaced by the thermometer.
>
> L.T. Wilkins: Social Deviance, page 9

With all the results so far given, the presentation has been on the basis of coordination level cut-offs. The reader is invited to consider the same test results, but now presented on the basis of a simulated ranking order and a document output cut-off. In Chapter 3, one of the main problems considered was that of totalling the results of a set of questions that was heterogenous in having different numbers of starting terms and matching terms. Several solutions were considered, but only brief mention was made of one possible method, namely document output cut-off. Although this method was recognised as having many advantages, it was decided not to use it for the main test results; this was partly because of the additional effort required to obtain the necessary prerequisite of a ranking order, but also because it would have involved a transformation of the test results as actually obtained by the co-ordination level cut-off. At a later date a simpler method of deriving a simulated ranking order was found and, in trying this out, it was shown that there was a possibility of obtaining an 'area measure' which could be used for producing an order of performance effectiveness for the different index languages. Therefore, the majority of the test searches were converted to a simulated ranking order, and in this chapter the results are presented by the document output cut-off method.

The influence of the SMART system was mainly responsible for our original investigation into attempting to obtain a ranked output for the Cranfield test searches. In the SMART system, the output of a search is arranged in an order of decreasing correlation with the search question; this is established by each document having a scoring that is obtained by calculations based on the match between the request terms and the document terms in the particular dictionary being tested. Thus every document in the collection is assigned a rank order number, the rank position reflecting the correlation with the search system. A sample output from the SMART system, showing the results for Question 147 searched on the Cranfield 200 document collection for fourteen different options, is given in Fig. 5.1. This output sheet shows, for each of the fourteen options, the file numbers of the fifteen highest ranked documents and also the rank numbers of the five documents which are relevant to this particular question. The heading at the top of each section refers to the particular option being tested, and it can be seen that, with 'ABSTR OLD QS', for instance, the five relevant documents, Nos. 708, 711, 713, 712 and 709 were ranked 21, 32, 68, 76 and 122 respectively.

In Fig. 5.2. are shown the conventional search results for 42 questions by Index Language I.1.a, and these are set out in coordination levels.

Q147CONTROLS 5 RELEVANT

	ABSTR OLD QS		ABSTR F NULL		ABSTR NEW QS		INDEX OLD QS		INDEX F NULL		INDEX NEW QS		CRAN CONCON		ABSTR F NULL INDEX F NULL	
	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT
1	792	21 708	792	13 712	792	7 712	792	18 712	792	14 712	792	3 712	597	4 708	792	20 712
2	33H	32 711	670	21 708	797	19 711	748	36 708	799	26 711	683	69 708	717	60 712	670	32 708
3	797	68 713	880	22 711	799	97 713	800	83 713	748	35 708	712	80 711	748	65 709	799	33 711
4	683	76 712	33H	41 713	324	101 708	437	117 711	316	147 709	797	139 713	708	98 713	880	56 713
5	574	122 709	33I	76 709	683	149 709	799	137 709	11A	166 713	988	153 709	451	128 711	748	99 709
6	799		874		992		797		683		572		37I		33H	
7	880		700		712		572		670		437		10A		316	
8	655		451		707		683		08D		799		916		33I	
9	995		704		416		07D		666		682		509		11A	
10	569		919		321		509		701		675		982		874	
11	800		748		655		988		33H		670		700		683	
12	793		988		33H		656		451		316		792		700	
13	416		712		874		10B		572		681		33I		451	
14	700		415		316		983		712		701		797		08D	
15	34A		08D		919		701		705		707		794		704	
RNK REC=	0.0470		0.0867		0.0402		0.0384		0.0387		0.0338		0.0423		0.0625	
LOG PRE=	0.2410		0.2859		0.2509		0.2327		0.2448		0.2433		0.2508		0.2577	
NOR REC=	0.6882		0.8379		0.6328		0.6144		0.6174		0.5600		0.6513		0.7692	
NOR PRE=	0.3029		0.4471		0.3390		0.2701		0.3172		0.3115		0.3386		0.3624	
OVERALL=	0.2880		0.3726		0.2911		0.2710		0.2835		0.2771		0.2930		0.3202	
NOR OVR=	0.3029		0.6369		0.3390		0.2701		0.3172		0.3115		0.3386		0.3624	

	ABSTR NEW QS INDEX NEW QS		ABSTR NEW QS ABSTR F NULL		INDEX NEW QS INDEX F NULL		CRAN CONCON ABSTR NEW QS		CRAN CONCON INDEX NEW QS		CRAN CONCON INDEX F NULL	
	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT
1	792	5 712	792	12 712	792	4 712	597	7 708	597	6 712	597	6 708
2	797	30 711	797	31 711	683	40 711	792	14 712	792	7 708	792	25 712
3	683	96 708	670	35 708	799	53 708	717	36 711	717	103 709	717	46 711
4	799	122 713	799	60 713	712	165 713	797	102 709	683	122 711	799	102 709
5	712	171 709	880	102 709	748	175 709	748	136 713	748	138 713	748	136 713
6	324		324		797		799		712		708	
7	988		33H		316		708		708		316	
8	992		683		988		324		797		451	
9	572		33I		11A		451		451		11A	
10	437		992		572		683		988		37I	
11	707		874		437		37I		37I		683	
12	416		712		670		992		572		10A	
13	682		700		08D		10A		10A		670	
14	321		707		682		712		437		916	
15	675		451		666		916		916		08D	
RNK REC=	0.0354		0.0625		0.0343		0.0508		0.0399		0.0476	
LOG PRE=	0.2453		0.2631		0.2478		0.2704		0.2644		0.2605	
NOR REC=	0.5805		0.7692		0.5672		0.7128		0.6297		0.6923	
NOR PRE=	0.3188		0.3802		0.3283		0.4027		0.3843		0.3717	
OVERALL=	0.2806		0.3256		0.2822		0.3212		0.3043		0.3081	
NOR OVR=	0.3188		0.3802		0.3283		0.4027		0.3843		0.3717	

FIGURE 5.1T EXAMPLE OF SMART OUTPUT WITH CRANFIELD 200 DOCUMENT COLLECTION (QUESTION 147)

	1+		2+		3+		4+		5+		6+		7+		8+		9+	
Q	R	N	R	N	R	N	R	N	R	N	R	N	R	N	R	N	R	N
79	2	60	1	7	1	1												
100	4	167	3	71	3	50	1	2										
116	6	169	5	92	4	51	3	25	1	12	0	4						
118	5	123	5	49	5	31	3	17	3	10	1	2	1	0				
119	6	170	6	82	4	37	3	13	3	6	0	1						
121	3	30	3	5	3	0	3	0	1	0								
122	5	107	5	41	3	11	1	0										
123	3	92	3	24	3	3												
126	2	95	2	62	2	20	2	4	2	0								
130	4	153	4	28	1	0												
132	2	78	2	51	1	31	1	5										
136	6	63	6	23	6	8	6	4	5	4	5	3	3	2	2	1	0	1
137	6	147	6	60	6	22	6	10	2	4								
141	1	82	1	35	1	4	1	1										
145	12	168	12	102	11	47	7	23	6	10	4	1						
146	7	37	4	1	1	1												
147	3	97	2	35	1	13	0	7	0	3	0	1						
148	4	36	4	15	4	4	2	0	1	0								
167	4	182	4	105	3	51	3	21	3	9	1	1						
170	2	109	1	45	1	18	1	6	1	1								
181	2	164	2	42	1	7												
182	3	175	1	47	1	5												
189	2	64	0	12	0	1												
190	7	162	6	45	5	10	3	0										
223	2	148	2	75	2	38	2	19	2	3	2	1	2	0				
224	5	50	4	65	2	27	0	3										
225	6	158	4	91	4	43	2	17	0	5								
226	7	60	4	19	4	2	4	1	2	0								
227	2	83	2	35	2	8	2	3	1	0	1	0						
230	7	42	2	0	1	0												
250	8	162	8	54	8	25	8	7	5	4	3	0						
261	4	131	4	34	4	13	4	5	4	0	4	0	3	0				
264	2	104	2	29	2	5	2	1	1	0	1	0	1	0				
266	5	164	4	32	0	8	0	1										
268	5	23	5	1	4	0	2	0										
269	4	34	4	4	2	0	1	0										
272	4	183	4	123	4	66	4	22	3	4	3	1	2	0	2	0		
273	7	33	6	10	5	1	2	0										
274	5	177	4	81	3	28	2	8										
317	2	118	2	69	2	31	2	10	0	2	0	1						
323	5	162	5	69	4	26	0	2	0	1								
360	8	143	8	59	8	14	5	4	3	2	0	1						

FIGURE 5.2T SEARCH RESULTS BY COORDINATION LEVEL CUTOFF FOR SINGLE TERM INDEX LANGUAGE (I.1.a) WITH 42 QUESTIONS AND 200 DOCUMENT COLLECTION.

(R = Relevant documents retrieved

N = Non-relevant documents retrieved)

By using these figures it was found possible to obtain a simulated ranking output. This is done by assigning a rank order number to each relevant document retrieved by means of the equations:-

$$^{c}R_{n} = X_{c} + (n - Y_{c})\left(\frac{x_{c} + 1}{y_{c} + 1}\right)$$

where $^{c}R_{n}$ is the rank order number of the n^{th} relevant document to be retrieved

- c is the coordination level at which the n^{th} relevant document is retrieved
- x_c is the additional number of documents retrieved at coordination level c. (i.e. those not retrieved at a higher coordination level)
- y_c is the additional number of relevant documents retrieved at coordination level c. (i.e. those not retrieved at a higher coordination level)
- X_c is the total number of documents retrieved before searching at coordination level c. (i.e. at higher coordination levels)
- Y_c is the total number of relevant documents retrieved before searching at coordination level c. (i.e. at higher coordination levels)

$^{c}R_{n}$ is taken to the nearest whole number but if its value falls exactly between two whole numbers, it is taken to the lower whole number for odd numbered questions and to the higher whole number for even numbered questions. Two examples to illustrate the effect are taken from Fig. 5.2. With Question 100, no documents are retrieved at a coordination level higher than four, so for this question, the various values are as follows:

Question 100

At level c=4, then $x_4 = 3$, $y_4 = 1$, $X_4 = 0$, $Y_4 = 0$

At level c=3, then $x_3 = 50$, $y_3 = 2$, $X_3 = 3$, $Y_3 = 1$

At level c=2, then $x_2 = 21$, $y_2 = 0$, $X_2 = 53$, $Y_2 = 3$

At level c=1, then $x_1 = 97$, $y_1 = 1$, $X_1 = 74$, $Y_1 = 3$

$\therefore$ For Relevant Document 1, retrieved at level 4 :-

$$^{4}R_{1} = 0 + (1 - 0)\left(\frac{3 + 1}{1 + 1}\right) = 0 + 2 = 2$$

For Relevant Document 2, retrieved at level 3 :-

$$^{3}R_{2} = 3 + (2 - 1)\left(\frac{50 + 1}{2 + 1}\right) = 3 + 17 = 20$$

For Relevant Document 3, retrieved at level 3 :-

$$^{3}R_{3} = 3 + (3 - 1)\left(\frac{50 + 1}{2 + 1}\right) = 3 + 34 = 37$$

For Relevant Document 4 retrieved at level 1 :-

$$^{1}R_{4} = 74 + (4 - 3)\left(\frac{97 + 1}{1 + 1}\right) = 74 + 49 = 123$$

In the next example considered, Question 123, there are actually four relevant documents; no documents are retrieved at a coordination

Q	REL	1	2	3	4	5	6 -7	8 -10	11 -15	16 -20	21 -30	31 -50	51 -75	76 -100	101 -125	126 -150	151 -175	176 -200
79	3	x									x					x		
100	4		x							x	x					x		
116	6							x		x	x	x	x		x			
118	5	x					x	x			xx							
119	6			x		x	x				x		xx					
121	3	x	x	x														
122	5	x				x		x			x	x						
123	4	x		x		x											x	
126	2	x	x															
130	4	x						x	x		x							
132	4				x							x			x		x	
136	6		x	x		x	x	xx										
137	6		x			x		xx	xx									
141	1	x																
145	12	x	x	x	x			x	x	x		xxx	x	x				
146	9		x	x	x	x			x		x	x		x		x		
147	5							x			x		x			x	x	
148	4	x	x		x		x											
167	4	x				x		x						x				
170	2		x											x				
181	2				x						x							
182	4				x									x	x			x
189	2										x	x						
190	7	x	x	x			x		x			x			x			
223	2	x	x															
224	5								x		x	x	x	x				
225	6							x	x		x	x			x	x		
226	7	x	x	x		x						xx	x					
227	2	x		x														
230	7	x	x					x		x	x	xx						
250	8	x	x	x		x	x	x	xx									
261	4	x	x	x	x													
264	2	x	x															
266	5								x	x	x	x		x				
268	5	x	x	x	x		x											
269	4	x	x		x		x											
272	4	x	x		x					x								
273	7	x	x	x	x		x		x		x							
274	5				x		x			x			x			x		
317	2					x		x										
323	5					x			x	x	x	x						
360	8		x		x	x	x	x	x	xx								
Totals		23	21	13	13	12	11	16	14	10	18	17	8	7	5	6	3	1
Recall		12	22	29	35	41	47	56	62	67	76	85	89	92	95	98	99	100
Precision		55	51	45	42	39	32	26	20	16	12	8	6	4	4	3	3	2

FIGURE 5.3T DOCUMENT OUTPUT CUT-OFF SCORE SHEET FOR INDEX LANGUAGE I.1.a FOR 42 QUESTIONS WITH 200 DOCUMENT COLLECTION.

level higher than three. It will be seen from Fig. 5.2. that at the single term level, only three of these documents have been found. The remaining relevant document can only be retrieved by searching through the remainder of the collection, namely 105 documents, and therefore at $c=0$, x_o is taken to be 105. In addition the equations do not always produce whole numbers, so $^{c}R_{n}$ has to be taken to the nearest whole number, or to the lower whole number where the value falls exactly between two whole numbers (since Q123 is an odd-numbered question).

<u>Question 123</u>

At level $c=3$, then $x_3 = 6$, $y_3 = 3$, $X_3 = 0$, $Y_3 = 0$

At level $c=2$, then $x_2 = 21$, $y_2 = 0$, $X_2 = 6$, $Y_2 = 3$

At level $c=1$, then $x_1 = 68$, $y_1 = 0$, $X_1 = 27$, $Y_1 = 3$

At level $c=0$, then $x_o = 105$, $y_o = 1$, $X_o = 95$, $Y_o = 3$

Then :-

$$^{3}R_{1} = 0 + (1 - 0)\left(\frac{6 + 1}{3 + 1}\right) = \frac{7}{4} = 2$$

$$^{3}R_{1} = 0 + (2 - 0)\left(\frac{6 + 1}{3 + 1}\right) = \frac{7}{2} = 3$$

$$^{3}R_{3} = 0 + (3 - 0)\left(\frac{6 + 1}{3 + 1}\right) = \frac{21}{4} = 5$$

$$^{o}R_{4} = 95 + (4 - 3)\left(\frac{105 + 1}{1 + 1}\right) = 95 + 53 = 148$$

The argument for this simulated ranking method is given in Appendix 5A.

When all such rankings have been calculated for the searches with a single index language, the results are entered on a score sheet as in Fig. 5.3T, which represents the results as given in Fig. 5.2T. Seventeen ranking groups were selected to have approximately the same number of documents falling in to each group; these were 1; 2; 3; 4; 5; 6-7; 8-10; 11-15; 16-20; 21-30; 31-50; 51-75; 76-100; 101-125; 126-150; 151-175; and 176-200. A cross is put in the appropriate column of the score sheet for every relevant document for the 42 questions. From the score-sheet, the total number of relevant documents retrieved at each of the seventeen cut-off levels can now be obtained. In Fig. 5.3T it is shown that, in the 42 searches, the first document retrieved was relevant on 23 occasions. As there were 198 documents relevant to the 42 questions, the recall ratio at this stage can be calculated as $\frac{23}{198} \times 100 = 12\%$; the precision ratio is calculated on the basis of one document having been retrieved for each question, and is therefore $\frac{23}{42} \times 100 = 55\%$. In 21 of the searches, the second document retrieved was relevant, making a total of 44 relevant documents so far retrieved, so the recall ratio increases to 22%. The precision ratio is now calculated on the basis of 2 x 42 documents having been retrieved, and is therefore 51%. Recall and precision ratios are similarly calculated for each document output cut-off level; ultimately the recall ratio will reach 100%.

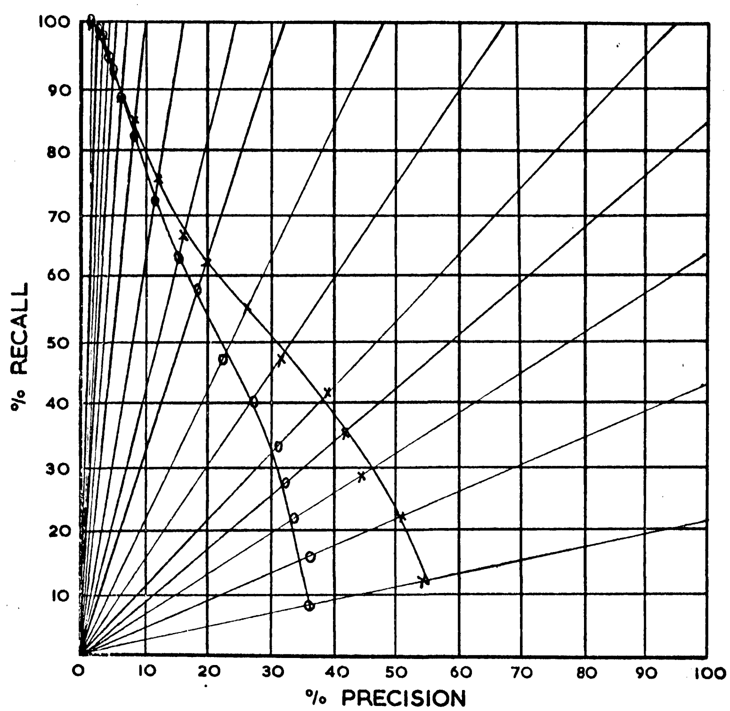

FIGURE 5.4P PLOT OF RESULTS FOR INDEX LANGUAGES I.1.a AND I.9.a BY DOCUMENTS OUTPUT CUT-OFF METHOD, SHOWING DOCUMENT OUTPUT CUT-OFF LINES.

x INDEX LANGUAGE I.1.a

o INDEX LANGUAGE I.9.a

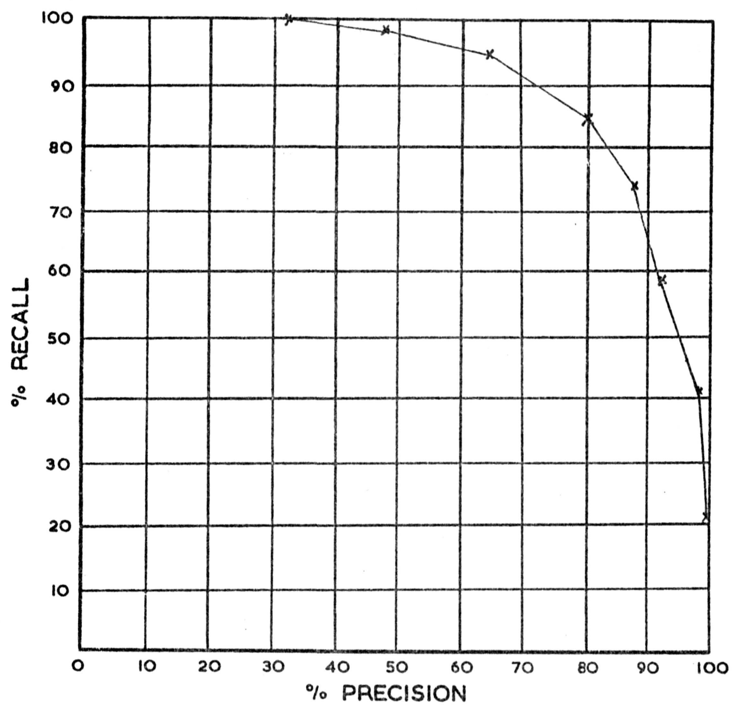

FIGURE 5.5P MAXIMUM POSSIBLE PERFORMANCE CURVE WITH DOCUMENT OUTPUT CUT-OFF FOR CRANFIELD TEST COLLECTION OF 200 DOCUMENTS AND 42 QUESTIONS.

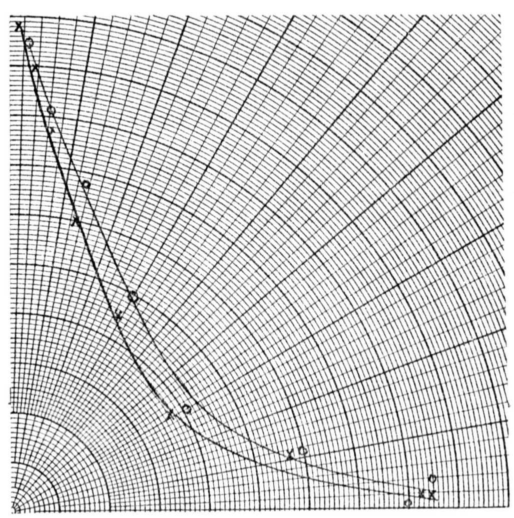

FIGURE 5.6P POLAR GRAPH OF RESULTS FOR INDEX LANGUAGE I.1.a (FIG. 4.140T) AND INDEX LANGUAGE I.6.a(FIG. 4.203T) (COORDINATION LEVEL CUTOFFS)

o I.1.a x I.6.a

Such recall and precision figures can be plotted on a conventional graph as in Fig. 5.4P, which shows the results of index language I.1.a (as in Fig. 5.3T) and also index language I.9.a. These curves can be compared with Fig. 4.206P and show the same superiority of index language I.1.a over index language I.9.a.

There is, however, an important difference. The positions of the points in Fig. 4.206P were determined by coordination level cut-offs, and were therefore random in relation to each other. With Fig. 5.4P, if straight lines are drawn radiating from the point of origin, these will, as can be seen, pass through the corresponding points in each curve. This is due to the fact that the cut-off is based on document output, and recall and precision ratios are now interdependent. It is known that there are 198 documents relevant to the 42 questions, so, on average, 4.7 documents are relevant to each question. When only one document is retrieved for each question, even if every such document were relevant, the recall ratio could not possibly be higher than $\frac{100 \times 42}{198} = 21.2\%$, although it would, of course, represent a precision ratio of 100%. If any of the documents are not relevant, then the recall ratio will always fall on some point along the line which goes from the point of origin to a recall of 21.2% at 100% precision. Therefore at any given document output cut-off, a drop in recall ratio with any one system as against any other system must also involve a drop in the precision ratio. Similarly, when two documents are retrieved in each search, the maximum recall ratio is 42.4% and with this particular document/question set, 100% recall cannot possibly be reached until at least five documents are retrieved for each question. This would, however, represent a total of 210 documents. Since there are only 198 relevant documents in the collection, the theoretical maximum precision ratio would then be $\frac{198}{210} \times 100 = 94.3\%$. As more documents are retrieved, so the maximum possible precision ratio must drop, and these document output cut-off performance lines can be calculated as has been done in Fig. 5.4P.

Because of the fact that Question 141 had only one relevant document, it would not be possible in this collection to obtain the theoretically maximum figures for recall and precision beyond the single document cut-off level. Similarly, there are thirteen questions which have more than five relevant documents, and 100% recall could not possibly be obtained until twelve documents have been retrieved, this number representing the highest figure for documents relevant to a single question. This does not affect the position of the lines, which would be different, however, for other situations where there are more or less relevant documents per question.

As previously mentioned, it is not possible to obtain the theoretically maximum performance beyond the single document output cut-off, since Q141 has only one relevant document. As ten questions have only two relevant documents, there must be a further deviation from the theoretical maximum beyond this stage. In Fig. 5.5P is shown the actual possible maximum performance that could be obtained with this collection. Achieving this performance would imply that for each question all the relevant documents were retrieved before any non-relevant documents were retrieved.

In Fig. 5.4P the lines radiating from the point of origin have been based on the document output cut-off for this particular test situation, but the performance curves could be drawn on a polar coordinate graph with the lines radiating at regular intervals as in Fig. 5.6P. The original purpose of using this type of graph was to investigate the possibility that

Q	REL	1	2	3	4	5	6 -7	8 -10	11 -15	16 -20	21 -30	31 -50	51 -75	76 -100	101 -125	126 -150	151 -175	176 -200
79	3											x	x	x				
100	4	x		x		x	x											
116	6			x				x	x	x	xx							
118	5				x					x	x	x	x					
119	6	x	x			x					x		x		x			
121	3	x	x	x														
122	5	x		x					x			x	x					
123	4		x		x			x				x						
126	2	x	x															
130	4								x			x			xx			
132	4		x		x				x						x			
136	6	x	x	x	x	x	x											
137	6			x				xx	x		x	x						
141	1	x																
145	12	x	x	x	x	x	x	x	x xx				x	x				
146	9		x					xx		x		xx	xx			x		
147	5						xx								xx	x		
148	4	x	x				x							x				
167	4							xx		x	x							
170	2		x										x					
181	2				x			x										
182	4										x			x		xx		
189	2		x	x														
190	7	x	x	x	x		x	x	x									
223	2	x	x															
224	5								x	x	x	x	x					
225	6							x			x	x	xx		x			
226	7	x	x				x		x	xx			x					
227	2	x	x															
230	7		x		x				xx	x	x		x					
250	8	x	x	x	x	x		x	x		x							
261	4	x	x	x	x													
264	2		x		x													
266	5						x		xx	x	x							
268	5	x	x		x			xx										
269	4	x	x	x					x									
272	4		x						xx	x								
273	7	x		x	x	x	xx	x										
274	5		x					x	x						x	x		
317	2							x	x									
323	5							x	x		x			xx				
360	8	x	x		x		x	x	x	x	x							
Totals		19	24	13	14	6	12	20	23	11	14	10	13	6	8	5	-	-

FIGURE 5.7T DOCUMENT OUTPUT CUTOFF SCORE SHEET FOR SMART 'CRAN CON-CON, INDEX NEW QS' FOR 42 DOCUMENTS WITH 200 DOCUMENT COLLECTION

comparison could be made between different index languages by measuring the performance over the whole curve, and the polar coordinate graphs were first tried with the performance curves obtained by the conventional coordination level cut-off as given in Chapter 4, where there was no direct relationship between the various cut-offs. The intention was to calculate the area encompassed by the performance curve within certain limits; with Fig. 5.6P (which is similar to Fig. 4203P) it was calculated that, in the area bounded by 95% recall and 85% precision, Index Language I.1.a had an area measure of 24.9 while Index Language I.6.a had an area measure of 21.1. It seemed to be unnecessary to do this with these new plots, since the document output cut-off automatically gave an exact match between systems. It was therefore hypothesised that obtaining a normalised recall ratio for all the systems tested would permit an 'order of effectiveness' to be determined. To obtain this normalised recall ratio, the recall ratio at each of the seventeen document cut-off levels would be summed and then divided by seventeen.

It was possible to test this idea by using the output from the SMART searches on the same collection. As previously stated, Professor Salton had results for fourteen different options, and Fig. 5.1T shows the output for question 147. Having similar output sheets for all 42 questions, it was possible to prepare a score sheet for each option. As an example the score sheet for 'Cran. Con Con Index News QS' is shown in Fig. 5.7T. Reference to Fig. 5.1T will show that the five relevant documents for Question 147 were ranked at 6, 7, 103, 122 and 138, and it can be seen that this is shown in the appropriate columns of Fig. 5.7T. The recall and precision ratios based on this procedure were obtained for the fourteen SMART options and the results are shown in Fig. 5.8T. The normalised recall ratios for each option were then calculated and are shown in Fig. 5.9T. A normalised recall and normalised precision for each question is given in the output sheets of the SMART searches (see Fig. 6.1) and finally calculated for the complete set of questions; the figures so obtained are also given in Fig.5.9T. In Fig.5.10T these two sets of results are arranged in order of effectiveness the higher figures representing the better results. It will be seen that, with very minor variations, the order obtained by the Cranfield normalised recall is the same as that obtained with the SMART normalised recall, with a rank correlation of +.991. This would appear to validate the ranking groups used at Cranfield, and also the simple method we have used to obtain the normalised recall ratio.

To sum up what has been so far discussed,the document ranking method has two major advantages.

1. It enables a series of cut-offs to be applied with equal consistency (i.e. an equal cut-off ratio, $\frac{100\,(a + b)}{N}$) between tests of different systems using the same document/question sets, and thus solves the problem of totalling sets of results which was discussed in Chapter 3.

2. It enables a series of recall ratios to be obtained which are directly comparable, and permits the calculation of a single measure of performance, normalised recall.

Regarding the measure itself, it was conceived (in a slightly different form) and originally used by Professor Salton. It is a method of representing performance over the whole of the operational range and therefore differs fundamentally from the 'single-point composite measures' which were discussed in Chapter 3. In experimental work of the nature described

DOCUMENTS OUTPUT CUT-OFF	S1		S2		S3		S4		S5		S6		S7		S8		S9		S10		S11		S12		S13		S14	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P
1	8	36	10	45	10	45	8	38	10	45	12	57	10	45	10	45	10	45	10	45	12	57	10	45	10	45	10	45
2	15	36	19	44	18	42	17	40	18	43	24	57	18	43	18	43	20	48	17	40	22	51	18	43	22	51	19	44
3	21	33	26	40	24	38	22	35	26	40	28	44	24	37	25	40	27	43	25	40	27	43	24	38	28	44	25	40
4	26	31	29	35	31	36	27	32	30	35	33	39	29	34	33	39	34	40	31	36	32	38	31	36	31	36	33	39
5	31	29	32	30	35	33	31	30	36	34	36	34	34	32	38	36	39	37	36	34	38	36	35	33	38	36	36	34
6-7	36	24	40	27	41	28	37	25	43	29	45	30	43	29	45	30	45	31	43	29	47	32	42	28	44	30	45	30
8-10	44	21	48	23	51	24	40	19	51	24	52	25	51	24	51	24	53	25	50	24	54	25	53	25	55	26	56	26
11-15	52	16	58	18	61	19	51	16	58	18	63	20	59	18	59	19	62	20	62	20	63	20	65	20	66	21	65	20
16-20	57	13	65	15	66	16	56	13	62	15	68	16	63	15	65	15	69	16	67	16	67	16	69	16	72	17	70	17
21-30	66	10	71	11	76	12	62	10	73	12	73	12	68	11	75	12	79	12	74	12	74	12	76	12	79	12	76	12
31-50	77	7	79	7	82	8	76	7	84	8	85	8	78	7	83	8	87	8	82	8	84	8	82	8	84	8	84	8
51-75	87	5	87	5	89	6	86	5	92	6	92	6	88	6	91	6	92	6	89	6	93	6	89	6	90	6	91	6
76-100	92	4	90	4	92	4	92	4	95	5	96	5	93	4	95	5	95	5	92	4	96	5	93	4	93	4	95	4
101-125	94	4	92	3	95	4	96	4	96	4	98	4	96	4	98	4	98	4	96	4	98	4	97	4	97	4	97	4
126-150	95	3	95	3	97	3	97	3	97	3	99	3	98	3	98	3	99	3	97	3	98	3	98	3	100	3	100	3
151-175	96	3	97	3	98	3	98	3	98	3	99	3	99	3	98	3	99	3	99	3	99	3	100	3	100	3	100	3
176-200	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2

S1 Abstracts old qs.
S2 Abstracts f null.
S3 Abstracts new qs.
S4 Indexing old qs.
S5 Indexing f null.
S6 Indexing new qs.
S7 Cranfield concon.
S8 Abstracts and indexing f null.
S9 Abstracts and indexing new qs.
S10 Abstracts new qs. and indexing f null.
S11 Indexing new qs and f null.
S12 Cranfield concon and abstracts new qs.
S13 Cranfield concon and indexing new qs.
S14 Cranfield concon and indexing f null.

FIG. 5.8T **RECALL AND PRECISION RATIOS FOR 14 SMART LANGUAGES BASED ON CRANFIELD DOCUMENT OUTPUT CUTOFF.**
(R = Recall Ratio, P = Precision Ratio)

SMART LANGUAGE	CRANFIELD NORMALISED RECALL	SMART NORMALISED RECALL AND PRECISION
S1	58.64	1.492
S2	61.06	1.546
S3	62.70	1.573
S4	58.58	1.495
S5	62.41	1.573
S6	64.88	1.609
S7	61.82	1.548
S8	63.64	1.594
S9	65.13	1.618
S10	62.94	1.579
S11	64.94	1.617
S12	63.64	1.593
S13	65.23	1.624
S14	64.82	1.612

FIGURE 5.9T PERFORMANCE FIGURES FOR SMART LANGUAGES

ORDER	CRANFIELD	SMART
1	S13	S13
2	S9	S9
3	S11	S11
4	S6	S14
5	S14	S6
6	S8	S8
7	S12	S12
8	S10	S10
9	S3	S3
10	S5	S5
11	S7	S7
12	S2	S2
13	S1	S4
14	S4	S1

FIGURE 5.10T COMPARISON OF RANKING OF SMART LANGUAGES BY CRANFIELD AND SMART NORMALISED MEASURES

in this report, it appears to give a valid single measure for comparing the performance of different systems, and, without wishing to be overdogmatic, appears more suitable for this purpose than anything else that has been proposed.

Having - to our satisfaction - established the reasonableness both of the simulated ranking method and also the method for obtaining normalised recall, the procedure was used for the four main groups of index languages. Fig. 5.11T gives the recall and precision ratios for the eight single term languages, while Fig. 5.12T gives similar figures for the fifteen concept languages. The results of the six controlled languages are given in Fig. 5.13T and the searches of titles and abstracts are shown in Fig. 5.14T. These tables also show the normalised recall ratio for each index language. In Fig. 5.15T the index languages are rearranged into an order based on this normalised recall ratio, from which it can be seen that the highest score (65.82) is obtained by Index Language I.3.a (single terms, word forms), with the lowest score (44.64) for Index Language II.1.a (single concepts, natural language). It will be noted that this table also includes the fourteen SMART options.

The figures given so far have been based on what has earlier been described as the average of numbers, and it might be thought that the document ranking method would be particularly susceptible to aberrations which the average of numbers sometimes produces. The results have therefore been recalculated by the average of ratios. To do this, as can be seen from the example in Fig. 5.16T, the indication of a relevant document is replaced by the number representing the percentage of the total recall ratio for that particular question. Thus, with question 79, there were three relevant documents, each document therefore representing 33.3% of the total. With question 100, having four relevant documents, each relevant document is 25% of the total. Question 141 has only one relevant document, so the retrieval of this single document represents 100% recall. These figures are summed for each column, then aggregated and finally, of course, reach a total of 4200. Recall figures can then be obtained.

This process was carried out for all the index languages, and as can be seen from Fig. 5.17T this results in a general increase of two or three points in the normalised recall ratio; however, when placed in order, as in Fig. 5.18T, it can be seen that this order is virtually unchanged from that obtained with the average of numbers, with a positive rank correlation of +.992.

Fig. 5.19T shows the result of ranking documents on the complete collection of 1400 documents. It covers the 42 questions with Index Language I.1.a., and is therefore directly comparable with Fig. 5.3T which was based on the smaller collection of 200 documents. The first eleven ranking groups have been retained, after which they are enlarged to take in the greater number of documents. Fig. 5.20P gives the performance curves for the two situations, and shows that, as would be expected, the smaller generality number for the 1400 document collection adversely affects the performance.

In Chapter 4, Section 8, were given the performance figures for the controlled term languages with Search E, which required some intellect to be applied to the search formulation. The result of ranking the output from these searches is given in Fig. 5.21T, and the

DOCUMENT OUTPUT CUT-OFF	I-1		I-2		I-3		I-5		I-6		I-7		I-8		I-9	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P
1	12	55	12	57	12	57	13	60	11	52	10	48	9	43	8	36
2	22	51	23	54	23	54	19	45	21	49	21	49	19	44	16	37
3	29	45	30	47	30	48	28	44	29	46	29	46	28	44	22	34
4	35	42	36	42	37	43	32	38	35	42	33	39	32	38	27	32
5	41	39	41	39	43	40	36	34	40	38	40	38	38	36	33	31
6-7	47	32	48	32	48	32	45	30	47	32	46	31	46	31	40	27
8-10	56	26	55	26	56	26	53	25	55	26	53	25	55	26	47	22
11-15	62	20	63	20	64	20	59	19	62	19	63	20	62	20	58	18
16-20	67	16	67	16	70	17	65	15	66	16	67	16	68	16	63	15
21-30	76	12	76	12	76	12	73	12	73	12	76	12	76	12	72	11
31-50	85	8	85	8	86	8	82	8	83	8	86	8	85	8	82	8
51-75	89	6	89	6	89	6	88	6	89	6	91	6	91	6	89	6
76-100	92	4	92	4	93	4	91	4	92	4	93	4	93	4	93	4
101-125	95	4	95	4	95	4	94	4	95	4	95	4	96	4	95	4
126-150	98	3	98	3	98	3	96	3	97	3	97	3	98	3	97	3
151-175	99	3	99	3	99	3	98	3	99	3	99	3	99	3	98	3
176-200	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2
NORMALISED RECALL	65.00		65.23		65.82		63.05		64.47		64.05		64.41		61.17	

SINGLE TERM LANGUAGES

I-1 Natural language
I-2 Synonyms
I-3 Word endings
I-5 Synonyms and quasi-synonyms
I-6 Synonyms, word endings, and quasi-synonyms
I-7 Hierarchical reduction first stage
I-8 Hierarchical reduction second stage
I-9 Hierarchical reduction third stage

FIGURE 5.11T RECALL AND PRECISION RATIOS AND NORMALISED RECALL FOR SINGLE TERM INDEX LANGUAGES (AVERAGE OF NUMBERS)
(R = Recall Ratio, P = Precision Ratio)

DOCUMENTS OUTPUT CUT-OFF	II - 1		II - 2		II - 3		II - 4		II - 5		II - 6		II - 7		II - 8		II - 9		II - 10		II - 11		II - 12		II - 13		II - 14		II - 15	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P
1	7	31	7	31	10	48	8	38	9	43	8	36	6	29	7	31	10	48	10	48	9	43	8	38	6	29	5	21	6	29
2	12	29	13	30	18	43	12	29	15	36	12	29	11	26	13	30	19	45	17	39	16	38	18	42	14	32	9	21	11	26
3	15	24	18	29	23	36	17	26	21	33	18	29	18	29	17	27	26	41	27	42	25	39	24	38	17	27	12	18	17	26
4	18	21	20	24	28	33	21	25	25	29	21	24	22	26	22	26	29	34	30	36	30	35	30	35	24	28	19	23	21	25
5	21	20	21	20	31	29	24	23	29	27	22	21	26	24	26	24	33	31	36	34	38	36	32	30	27	26	21	20	25	23
6-7	24	16	28	19	35	23	29	20	35	23	31	21	30	20	32	21	41	28	43	29	44	30	36	24	32	22	27	18	33	22
8-10	29	14	32	15	41	20	37	17	39	19	36	17	39	18	39	18	45	21	50	24	51	24	41	20	39	19	33	16	41	19
11-15	31	10	36	11	45	14	42	13	47	15	43	13	48	15	47	15	52	16	59	18	60	19	48	15	48	15	43	13	50	16
16-20	34	8	39	9	49	12	48	11	53	13	48	11	55	13	56	13	56	13	64	15	66	15	52	12	54	13	47	11	55	13
21-30	37	6	43	7	54	8	57	9	60	9	56	9	60	9	66	10	60	9	75	12	76	12	59	9	64	10	57	9	66	10
31-50	46	4	52	5	60	6	65	6	68	6	62	6	67	6	74	7	67	6	84	8	83	8	65	6	74	7	70	7	80	8
51-75	55	3	59	4	67	4	71	4	74	5	70	4	74	5	80	5	74	5	89	6	88	6	73	5	81	5	79	5	88	6
76-100	67	3	70	3	75	4	77	4	80	4	78	4	81	4	85	4	80	4	93	4	92	4	78	4	85	4	84	4	91	4
101-125	77	3	80	3	84	3	85	3	88	3	85	3	87	3	90	3	87	3	95	4	96	4	87	3	91	3	91	3	95	4
126-150	87	3	89	3	91	3	93	3	94	3	92	3	93	3	94	3	93	3	97	3	98	3	92	3	95	3	96	3	98	3
151-175	99	3	99	3	99	3	99	3	99	3	99	3	99	3	100	3	99	3	100	3	100	3	99	3	99	3	99	3	99	3
176-200	100	2	100	3	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2	100	2
NORMALISED RECALL	44.64		47.41		53.52		52.05		55.05		51.82		53.88		55.76		57.11		62.88		63.05		55.41		55.88		52.47		57.41	

SIMPLE CONCEPT INDEX LANGUAGES

II-1	Natural language
II-2	Synonyms
II-3	II-2 + species (selection)
II-4	II-2 + superordinate
II-5	II-3 + II-4
II-6	II-2 + coordinate (selection)
II-7	II-6 + collateral (selection)
II-8	II-5 + II-7
II-9	II-2 + alphabetical collateral first stage (selection)
II-10	II-9 + alphabetical collateral second stage (selection)
II-11	II-8 + II-10
II-12	II-2 + complete species
II-13	II-12 + superordinate
II-14	II-13 + complete collateral
II-15	II-11 + II-14

FIGURE 5.12T RECALL AND PRECISION RATIOS AND NORMALISED RECALL FOR SIMPLE CONCEPT INDEX LANGUAGES (AVERAGE OF NUMBERS)

(R = Recall Ratio, P = Precision Ratio)

DOCUMENT OUTPUT CUT-OFF	III-1 R	III-1 P	III-2 R	III-2 P	III-3 R	III-3 P	III-4 R	III-4 P	III-5 R	III-5 P	III-6 R	III-6 P
1	12	55	12	57	11	52	10	48	10	45	8	38
2	17	40	19	44	16	38	16	38	13	31	12	29
3	26	40	25	40	24	37	23	36	19	30	18	28
4	31	36	32	38	28	33	28	33	25	29	24	28
5	36	34	35	33	34	32	33	31	27	25	27	25
6-7	43	29	43	29	41	28	42	28	36	24	34	23
8-10	50	24	49	23	47	22	46	22	46	22	42	20
11-15	58	18	58	18	55	17	54	17	52	16	53	17
16-20	61	14	62	15	60	14	60	14	59	14	60	14
21-30	71	11	69	11	68	11	65	10	68	11	69	11
31-50	80	8	81	8	77	7	76	7	79	7	81	8
51-75	85	5	85	5	86	5	85	5	88	6	88	6
76-100	88	4	90	4	88	4	89	4	93	4	93	4
101-125	95	4	94	4	93	4	94	4	98	4	98	4
126-150	97	3	97	3	95	3	95	3	100	3	99	3
151-175	100	3	99	3	99	3	99	3	100	3	100	3
176-200	100	2	100	2	100	2	100	2	100	2	100	2
NORMALISED RECALL	61.76		61.76		60.11		59.70		59.58		59.17	

CONTROLLED TERM INDEX LANGUAGES

III-1 Controlled terms
III-2 Narrower terms
III-3 Broader terms
III-4 Related terms
III-5 Narrower and broader terms
III-6 III-4 + III-5

FIGURE 5.13T RECALL AND PRECISION RATIOS AND NORMALISED RECALL FOR CONTROLLED TERM INDEX LANGUAGES (AVERAGE OF NUMBERS)
(R = Recall Ratio, P = Precision Ratio)

DOCUMENT OUTPUT CUT-OFF	IV-1 R	IV-1 P	IV-2 R	IV-2 P	IV-3 R	IV-3 P	IV-4 R	IV-4 P
1	10	48	10	48	10	48	11	52
2	17	38	17	39	18	43	18	43
3	25	39	24	38	23	39	24	38
4	29	34	28	34	29	34	30	35
5	32	30	32	30	32	30	35	32
6-7	39	25	41	28	39	25	39	25
8-10	45	21	47	22	47	22	46	21
11-15	55	17	55	17	56	17	54	17
16-20	60	14	62	15	64	15	58	13
21-30	64	10	67	10	70	11	69	11
31-50	76	7	78	7	80	8	80	8
51-75	82	5	83	5	87	6	86	6
76-100	84	4	85	4	91	4	92	4
101-125	89	3	91	3	94	4	95	4
126-150	95	3	96	3	97	3	97	3
151-175	100	3	100	3	99	3	100	3
176-200	100	2	100	2	100	2	100	2
NORMALISED RECALL	58.94		59.76		60.94		60.82	

ABSTRACTS AND TITLES

IV-1 Titles natural language
IV-2 Titles word forms
IV-3 Abstracts and titles natural languages
IV-4 Abstracts and titles word forms

FIGURE 5.14T RECALL AND PRECISION RATIOS AND NORMALISED RECALL FOR SINGLE TERM SEARCHES ON ABSTRACTS AND TITLES (AVERAGE OF NUMBERS)
(R = Recall Ratio, P = Precision Ratio)

ORDER	NORMALISED RECALL	INDEXING LANGUAGE	
1	65.82	I-3	Single terms. Word forms
2=	65.23	I-2	Single terms. Synonyms
2=	65.23	S-13	SMART Concon and indexing new qs.
4	65.13	S-9	SMART Abstract and indexing new qs.
5	65.00	I-1	Single terms. Natural language
6	64.94	S-11	SMART Indexing new qs. and f null
7	64.88	S-6	SMART Indexing new qs.
8	64.82	S-14	SMART Concon and indexing f null
9	64.47	I-6	Single terms. Synonyms, word forms, quasi-synonyms
10	64.41	I-8	Single terms. Hierarchy second stage
11	64.05	I-7	Single terms. Hierarchy first stage
12	63.64	S-8	SMART Abstracts and indexing f null
12=	63.64	S-12	SMART Indexing new qs. and f null
14	63.05	I-5	Single terms. Synonyms. Quasi-synonyms
14=	63.05	II-11	Simple concepts. Hierarchical and alphabetical selection
16	62.94	S-10	SMART Abstracts new qs. and indexing f null
17	62.88	II-10	Simple concepts. Alphabetical second stage selection
18	62.70	S-3	SMART Abstracts new qs.
19	62.41	S-5	SMART Indexing f null
20	61.82	S-7	SMART Concon
21	61.76	III-1	Controlled terms
21=	61.76	III-2	Controlled terms. Narrower terms
23	61.17	I-9	Single terms. Hierarchy third stage
24	61.06	S-2	SMART Abstracts f null
25	60.94	IV-3	Abstracts. Natural language
26	60.82	IV-4	Abstracts. Word forms
27	60.11	III-3	Controlled terms. Broader terms
28	59.76	IV-2	Titles. Word forms
29	59.70	III-4	Controlled terms. Related terms
30	59.58	III-5	Controlled terms. Narrower and broader terms
31	59.17	III-6	Controlled terms. Narrower, broader and related terms
32	58.94	IV-1	Titles. Natural language
33	58.64	S-1	SMART Abstracts old qs.
34	58.58	S-4	SMART Indexing old qs.
35	57.41	II-15	Simple concepts. Complete combination
36	57.11	II-9	Simple concepts. Alphabetical first stage selection
37	55.88	II-13	Simple concepts. Complete species and superordinate
38	55.76	II-8	Simple concepts. Hierarchical selection
39	55.41	II-12	Simple concepts. Complete species
40	55.05	II-5	Simple concepts. Selected species and super ordinate
41	53.88	II-7	Simple concepts. Selected coordinate and collateral
42	53.52	II-3	Simple concepts. Selected species
43	52.47	II-14	Simple concepts. Complete collateral
44	52.05	II-4	Simple concepts. Superordinate
45	51.82	II-6	Simple concepts. Selected coordinate
46	47.41	II-2	Simple concepts. Synonyms
47	44.64	II-1	Simple concepts. Natural language

FIGURE 5.15T ORDER OF EFFECTIVENESS BASED ON NORMALISED RECALL FOR 33 CRANFIELD AND 14 SMART INDEX LANGUAGES (AVERAGE OF NUMBERS)

Q	REL	1	2	3	4	5	6 -7	8 -10	11 -15	16 -20	21 -30	31 -50	51 -75	76 -100	101 -125	126 -150	151 -175	176 -200
79	3	33										34				33		
100	4		25							25	25					25		
116	6							16		17	16	17	16		17			
118	5	20					20	20			40							
119	6			16		17	16				17		34					
121	3	33	34	33														
122	5	20				20		20			20	20						
123	4	25		25		25											25	
126	2	50	50															
130	4	25						25	25		25							
132	4				25							25			25		25	
136	6		16	17		16	17	34										
137	6		16			17		33	34									
141	1	100																
145	12	8	9	8	9			8	9	8		25	9	8				
146	9		11	11	11	11			11		11	11		11		12		
147	5							20			20		20			20	20	
148	4	25	25		25		25											
167	4	25				25		25						25				
170	2				50									50				
181	2		50								50							
182	4				25									25	25			25
189	2										50	50						
190	7	14	14	14			15		14			14			15			
223	2	50	50															
224	5								20		20	20	20	20				
225	6							16	17		16	17			17	17		
226	7	14	14	15		14						28	15					
227	2	50		50														
230	7	14	14					15		14	14	29						
250	8	12	13	12		13	13	12	25									
261	4	25	25	25	25													
264	2	50	50															
266	5								20	20	20	20		20				
268	5	20	20	20	20		20											
269	4	25	25		25		25											
272	4	25	25		25					25								
273	7	14	14	14	15		14				14			15				
274	5				20		20			20			20			20		
317	2					50		50										
323	5					20			20	20	20	20						
360	8		12		13	12	13	12	13	25								
Totals		677	512	260	288	240	198	306	208	174	378	330	134	174	99	127	70	25

FIGURE 5.16T DOCUMENT OUTPUT CUT-OFF SCORE SHEET AS FIGURE 5.3T CONVERTED TO AVERAGE OF RATIOS.

Index Language	Normalised Recall	Index Language	Normalised Recall
I-1	67.2	III-1	64.2
I-2	67.7	III-2	64.5
I-3	68.5	III-3	62.6
I-5	65.6	III-4	62.4
I-6	66.9	III-5	61.7
I-7	67.4	III-6	61.7
I-8	67.1		
I-9	63.5		
II-1	45.6	IV-1	61.5
II-2	49.0	IV-2	62.4
II-3	55.2	IV-3	62.7
II-4	53,5	IV-4	63.1
II-5	56.3		
II-6	53.8		
II-7	55.6		
II-8	56.8		
II-9	59.3		
II-10	64.9		
II-11	65.1		
II-12	57.2		
II-13	58.4		
II-14	55.0		
II-15	59.8		

FIGURE 5.17T NORMALISED RECALL FOR CRANFIELD INDEX LANGUAGES BASED ON AVERAGE OF RATIOS.

ORDER	INDEX LANGUAGE		ORDER	INDEX LANGUAGE	
1	I-3	(1)	25	IV-4	(26)
2	S-13	(2)	26	IV-3	(25)
3	I-2	(2)	27	III-3	(27)
4	I-7	(11)	28	III-4	(29)
5	S-11	(6)	29	IV-2	(28)
6	I-1	(5)	30	III-5	(30)
7	S-14	(8)	31	III-6	(31)
8	S-9	(4)	32	IV-1	(32)
9	S-8	(10)	33	S-1	(33)
10	I-6	(9)	34	S-4	(34)
11	S-6	(7)	35	II-15	(35)
12	S-8	(12)	36	II-9	(36)
13	S-12	(13)	37	II-13	(37)
14	S-10	(16)	38	II-12	(39)
15	I-5	(14)	39	II-8	(38)
16	II-11	(15)	40	II-5	(40)
17	S-3	(18)	41	II-7	(41)
18	II-10	(17)	42	II-3	(42)
19	III-2	(22)	43	II-14	(43)
20	S-5	(19)	44	II-6	(45)
21	S-2	(24)	45	II-4	(44)
22	III-1	(21)	46	II-2	(46)
23	I-9	(23)	47	II-1	(47)
24	S-7	(20)			

FIGURE 5.18T ORDER OF EFFICIENCY BASED ON NORMALISED RECALL FOR CRANFIELD AND SMART INDEX LANGUAGES CALCULATED BY AVERAGE OF RATIOS (FIGURES IN BRACKETS REPRESENT ORDER WHEN CALCULATED BY AVERAGE OF NUMBERS AS IN FIG. 5.14T)

Q	REL	1	2	3	4	5	6 -7	8 -10	11 -15	16 -20	21 -30	31 -50	51 -100	101 -200	201 -400	401 -600	601 -800	801 -1100	1101 -1400
79	3			x											x			x	
100	4										x			x	x			x	
116	6										x	x	x	x	x			x	
118	5		x							xx				xx					
119	6							x	x	x			x	x	x				
121	3	x	x	x															
122	5					x						x	x	x	x				
123	4						x		x	x								x	
126	2	x	x																
130	4	x											x	x	x				
132	4							x						x				x	x
136	6		x	x		x	xx		x										
137	6			x			x			x	x	xx							
141	1			x															
145	12	x		x		x	xx			x	x		x	xxxx					
146	9	x			x				x	x	x		x	xx				x	
147	5						x						x		x			x	x
148	4	x				x			x	x									
167	4						x			x		x			x				
170	2		x													x			
181	2								x					x					
182	4										x							xx	x
189	2													x	x				
190	7		x	x		x					x	x		x				x	
223	2	x		x															
224	5											x	xxx					x	
225	6										x	x		xx			x	x	
226	7	x	x			x		x							xxx				
227	2	x						x											
230	7	x					x						xx	xx	x				
250	8	x		x	x		x	x			xx	x							
261	4	x	x	x	x														
264	2	x					x												
266	5											x	xx	x		x			
268	5	x	x		x	x			x										
269	4	x		x					x		x								
272	4	x	x			x							x						
273	7	x	x	x		x	x		x				x						
274	5						x		x				x		x			x	
317	2											x	x						
323	5										x	x	xx	x					
360	8				x		x	x			x	x	x	xx					
Totals		17	11	12	5	9	14	6	10	9	13	13	21	25	14	2	1	13	3

FIGURE 5.19T DOCUMENT OUTPUT CUT-OFF SCORE SHEET FOR INDEX LANGUAGE I.1.a FOR 42 QUESTIONS WITH 1,400 DOCUMENT COLLECTION.

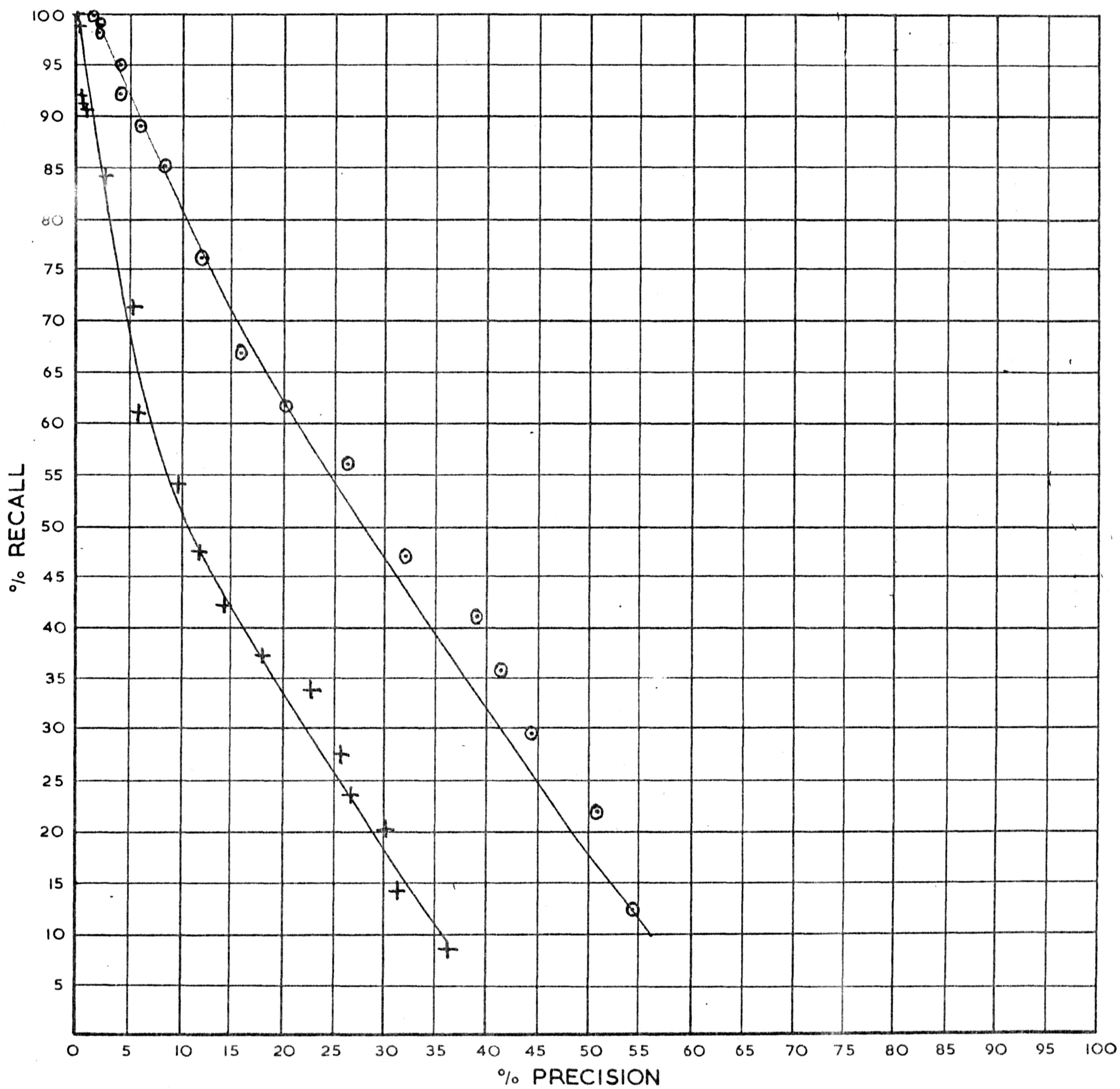

FIGURE 5.20P COMPARISON OF PERFORMANCE, BASED ON DOCUMENT OUTPUT CUT-OFF, FOR COLLECTIONS OF 1,400 AND 200 DOCUMENTS WITH INDEX LANGUAGE I.1.a AND 42 QUESTIONS.

\+ 1,400 Document collection. ⊙ 200 Document Collection normalised recall ratio is shown for each index language by Search E and Search A. It will be seen that there is an improvement with each language of from 1 to 2 points.

Fig. 5.22T shows the ranking score sheet for Index Language I.1.a. with the 42 questions on the 200 document collection, but with the lowest level of exhaustivity of indexing. Fig. 5.23P compares these results with those obtained under similar conditions except that exhaustivity was at its highest level (as Fig. 5.3T).

Four grades of document relevance were used in the tests, and the effect on performance of each of the relevance grades has been considered in Section 6 of Chapter 4. An alternative method of scoring performance from that so far used would be to take account of these relevance gradings by giving each document a weighting related to its relevance grading. The use of the document output cut-off method and normalised recall permits this to be done in what might be considered to be a meaningful manner. A simple form of weighting is to give a score of 4 to those documents rated relevance 1, a score of 3 for documents of relevance 2, a score of 2 for documents of relevance 3 and a score of 1 for documents rated relevance 4. The effect of this would be that question 119, for instance, which has two documents (1378 and 1667) rated relevance 2 and four documents (1324,1666, 1670 and 2391) rated relevance 3 would now have a total "retrieval score" of (2 x 3) + (4 x 2) = 14.

Referring back to Fig. 5.3T, the score sheet for this question would be amended to show the weighting of each relevant document according to the order in which the documents of the two levels of relevance were retrieved. This was done for the 42 questions by Index Language I.1.a and the amended score sheet is given as Fig. 5.24T. The recall ratio is now determined on the total "points" score for the set of questions, which is 421. At a document cut-off of 1, the recall ratio is therefore shown to be $\frac{58 \times 100}{421} = 14\%$ and the recall ratios are similarly calculated for the other sixteen cut-off groups. The normalised recall ratio is then calculated as being 67.12.

This procedure was repeated for five other index languages to find whether the effect of a weighting score made any difference to their comparative performance. As can be seen from Fig. 5.25T, there was for each case an increase of approximately two points in the normalised recall, so it does not appear that this method of weighting makes any significant difference to the overall comparison.

The exercise was repeated using different weightings, with a score of 10 for documents rated relevance 1, a score of 5 for documents rated relevance 2, a score of 3 for documents rated relevance 3 and a score of 1 for documents rated relevance 4. This resulted in a further small increase in the normalised recall ratios, but made no significant difference in the comparison between systems. It would be incorrect to state that some form of weighting might not be useful in certain circumstances, but it would seem that it does not have any particular value in this test.

In connection with the normalised recall ratio, it is obvious that there is what could be considered a minimum figure which is based on the random retrieval of the whole collection for every question. For instance, the three relevant documents of Question 79 would, with random retrieval, be ranked 50, 100 and 150, while the seven relevant documents of Question 190 would be ranked 25, 50, 75, 100, 125, 150 and 175. With this particular document/question set, the normalised recall ratio based on this random

DOCUMENT OUTPUT CUT-OFF	III-1		III-2		III-3		III-4		III-5		III-6	
	R	P	R	P	R	P	R	P	R	P	R	P
1	13	62	13	62	11	50	10	48	10	48	8	38
2	21	50	21	50	16	37	16	37	15	36	13	30
3	31	48	31	48	24	38	25	40	23	36	22	34
4	36	42	35	42	29	36	29	36	28	33	28	33
5	39	37	38	37	35	33	35	33	31	29	33	31
6-7	46	31	45	31	43	29	43	29	38	26	39	27
8-10	53	25	52	25	49	23	51	24	46	22	46	22
11-15	62	20	63	20	59	18	60	19	54	17	55	17
16-20	67	16	68	16	63	14	65	15	62	14	65	15
21-30	72	11	74	12	70	10	71	11	70	10	74	12
31-50	79	7	79	7	78	7	79	7	80	7	82	8
51-75	85	5	86	5	85	5	86	5	87	5	89	6
76-100	88	4	88	4	89	4	90	4	91	4	93	4
101-125	93	3	93	3	92	3	93	3	95	3	95	3
126-150	96	3	95	3	95	3	95	3	98	3	98	3
151-175	100	2	100	2	99	2	99	2	99	2	100	2
176-200	100	2	100	2	100	2	100	2	100	2	100	2
NORMALISED RECALL												
Search E	63.58		63.64		61.00		61.58		60.41		61.17	
Search A	61.76		61.76		60.11		59.70		59.58		59.17	

FIGURE 5.21T RECALL AND PRECISION FIGURES FOR INDEX LANGUAGES III.1 - III.6 FOR SEARCH E BY DOCUMENT OUTPUT CUT-OFF METHOD, TOGETHER WITH NORMALISED RECALL FOR SEARCH E AND SEARCH A.
(R = Recall Ratio, P = Precision Ratio)

Q	REL	1	2	3	4	5	6 -7	8 -10	11 -15	16 -20	21 -30	31 -50	51 -75	76 -100	101 -125	126 -150	151 -175	176 -200
79	3									x				x		x		
100	4							x			x		xx					
116	6			x					x	x		x	x			x		
118	5			x			x	x	x			x						
119	6	x					x	x	x		x		x					
121	3	x	x		x													
122	5		x		X			x	x			x						
123	4	x	x					x							x			
126	2	x									x							
130	4			x			x		xx									
132	4						x				x				x		x	
136	6	x	x	x		x	x	x										
137	6		x		x		x	x	x		x							
141	1	x																
145	12	x	x	x		x		x	x	x	xx	x		x		x		
146	9					x		x	x	xx			x	x		x	x	
147	5		x								x	x		x		x		
148	4		x	x										x		x		
167	4	x		x								x		x				
170	2	x										x						
181	2		x						x									
182	4		x										x			x	x	
189	2									x					x			
190	7	x	x		x		x	x		x			x					
223	2	x	x															
224	5						x					xx	xx					
225	6					x		x	x		x		x			x		
226	7	x	x	x	x						x	x			x			
227	2	x		x														
230	7	x					x	x	x	x	x				x			
250	8	x		x	x		xx	xx								x		
261	6	x	x	x	x													
264	2	x		x														
266	5							x			x	x	x			x		
268	5	x	x	x	x		x											
269	4		x	x		x	x											
272	4		x				x	x		x								
273	7	x	x	x	x			x	x	x								
274	5			x				x	x			x		x				
317	2			x								x						
323	5						x		x	x		x	x					
360	8			x		x	x	x		x	x	x	x					
Totals		19	18	18	9	6	16	19	15	12	13	15	13	7	5	10	3	

FIGURE 5.22T DOCUMENT OUTPUT CUT-OFF SCORE SHEET ON INDEX LANGUAGE I.1.a FOR 42 QUESTIONS WITH 200 DOCUMENT COLLECTION FOR INDEXING AT LEVEL OF EXHAUSTIVITY 1.

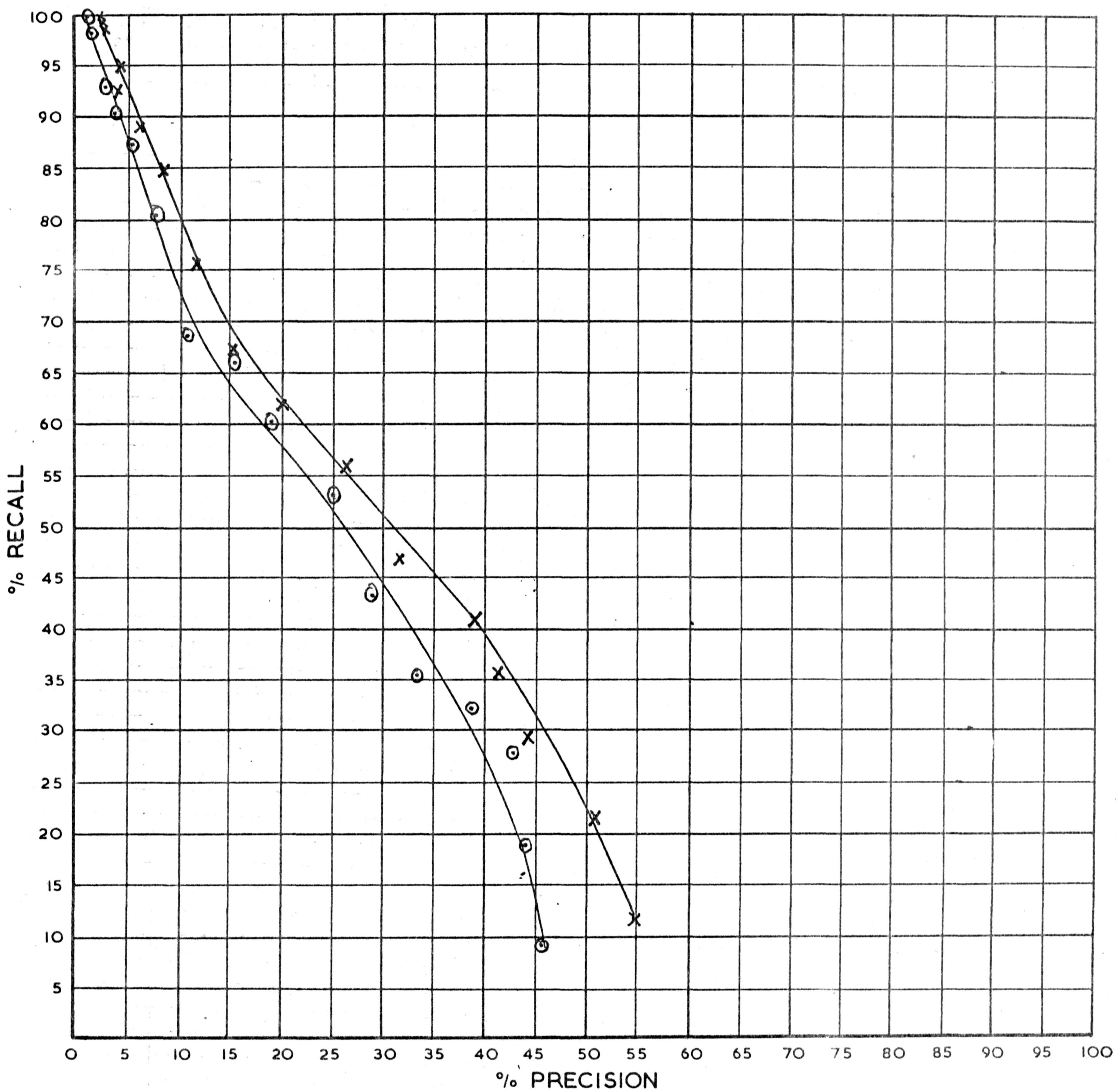

FIGURE 5.23P COMPARISON OF PERFORMANCE, BASED ON DOCUMENT OUTPUT CUT-OFF, FOR EXHAUSTIVITY LEVEL 1 (Fig. 5.22T) AND EXHAUSTIVITY LEVEL 3 (Fig. 5.3T).
o Exhaustivity Level 1 x Exhaustivity Level 3

Q	REL	1	2	3	4	5	6 -7	8 -10	11 -15	16 -20	21 -30	31 -50	51 -75	76 -100	101 -125	126 -150	151 -175	176 -200
79	5	2									2					1		
100	9		3							2	2					2		
116	11							3		2	3	1	1		1			
118	11	3					2	2			4							
119	14			2		3	2				2		5					
121	7	2	3	2														
122	7	2				1		1			1	2						
123	12	4		2		4											2	
126	5	4	1															
130	11	3						3	2		3							
132	11				2							4			2		3	
136	14		2	3		2	3	4										
137	11		3			2		4	2									
141	2	2																
145	16	2	1	2	1			2	1	1		4	1	1				
146	16		2	2	2	2			2		1	2		1		2		
147	7							1			1		1			1	3	
148	10	3	3		2		2											
167	11	4				3		2						2				
170	5		3											2				
181	5				2						3							
182	10				3									2	2			3
189	4										2	2						
190	15	2	3	2			2		3			2			1			
223	4	2	2															
224	10								2		3	1	1	3				
225	13							2	2		4	2			2	1		
226	11	1	1	2		2						4	1					
227	5	2		3														
230	17	2	2					3		2	4	4						
250	15	2	3	2		2	1	2	3									
261	12	4	4	2	2													
264	8	4	4															
266	13								3	2	3	2		3				
268	11	1	3	2	3		2											
269	8	2	1		4		1											
272	8	2	2		2					2								
273	12	3	1	2	1		2		2		1							
274	12				2		2			3			2			3		
317	6					3		3										
323	12					3			3	1	3	2						
360	15		2		2	2	2	2	1	4								
Totals	421	58	49	28	28	29	21	34	26	19	42	32	12	14	8	10	8	3

FIGURE 5.24T RESULTS AS Fig. 5.3T ADJUSTED FOR WEIGHTING BASED ON RELEVANCE GRADES.

Index Language	Normalised Recall Ratio (basic)	Normalised Recall Ratio (weighted)
I.1.a	65.00	67.12
I.7.a	64.05	65.94
III.1.a	61.76	63.64
III.6.a	59.17	61.06
II.9.a	57.11	58.94
II.5.a	55.05	57.11

FIGURE 5.25T COMPARISON OF NORMALISED RECALL RATIOS BY BASIC SCORING METHOD (as Fig. 5.15T) AND BY WEIGHTED SCORING METHOD FOR SIX INDEX LANGUAGES.

retrieval would be 26%. On the other hand, as was discussed earlier in this chapter, the theoretical maximum performance cannot be achieved due to the different numbers of relevant documents for each question, so the highest possible normalised recall ratio would be 86.70%.

It should also be emphasised that the normalised recall ratio only has meaning within the context of the manner in which it has been calculated. In this particular case it was by averaging the results of seventeen cut-off groups as given on page 198. Assume that the number of groups had been reduced to thirteen by combining the first six groups into two larger groups covering documents ranked 51 - 100 and documents ranked 101 - 200. The effect of doing this would be to reduce the normalised recall ratio for index language I.1.a from 65% to 55.7%. On the other hand, if the original groups were broken down so that no groups had more than ten rankings, the normalised recall ratio based upon the resulting twenty-seven groups would be 75.1%. At the same time, the effect of either of these actions would be to change, as considered in the previous paragraph, the minimum figure based on random retrieval and the maximum possible figure.